NCHRP
SYNTHESIS 300

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Performance Measures for Research, Development, and Technology Programs

A Synthesis of Highway Practice

TRANSPORTATION RESEARCH BOARD

NATIONAL RESEARCH COUNCIL

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

NCHRP SYNTHESIS 300

Performance Measures for Research, Development, and Technology Programs

A Synthesis of Highway Practice

CONSULTANT
SCOTT A. SABOL
Vermont Technical College
Randolph Center, Vermont

SUBJECT AREAS
Highway and Facility Design and Highway Operations, Capacity, and Traffic Control

Research Sponsored by the American Association of State Highway and Transportation Officials
in Cooperation with the Federal Highway Administration

TRANSPORTATION RESEARCH BOARD — NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. — 2001

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

In recognition of these needs, the highway administrators of the American Association of State Highway and Transportation Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by funds from participating member states of the Association and it receives the full cooperation and support of the Federal Highway Administration, United States Department of Transportation.

The Transportation Research Board of the National Research Council was requested by the Association to administer the research program because of the Board's recognized objectivity and understanding of modern research practices. The Board is uniquely suited for this purpose as it maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; it possesses avenues of communication and cooperation with federal, state, and local governmental agencies, universities, and industry; its relationship to the National Research Council is an insurance of objectivity; it maintains a full-time research correlation staff of specialists in highway transportation matters to bring the findings of research directly to those who are in a position to use them.

The program is developed on the basis of research needs identified by chief administrators of the highway and transportation departments and by committees of AASHTO. Each year, specific areas of research needs to be included in the program are proposed to the National Research Council and the Board by the American Association of State Highway and Transportation Officials. Research projects to fulfill these needs are defined by the Board, and qualified research agencies are selected from those that have submitted proposals. Administration and surveillance of research contracts are the responsibilities of the National Research Council and the Transportation Research Board.

The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs.

NOTE: The Transportation Research Board, the National Research Council, the Federal Highway Administration, the American Association of State Highway and Transportation Officials, and the individual states participating in the National Cooperative Highway Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

NCHRP SYNTHESIS 300

Project 20-5 FY 1999 (Topic 31-04)
ISSN 0547-5570
ISBN 0-309-06915-7
Library of Congress Control No. 2001 136162

Price $33.00

NOTICE

The project that is the subject of this report was a part of the National Cooperative Highway Research Program conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council. Such approval reflects the Governing Board's judgment that the program concerned is of national importance and appropriate with respect to both the purposes and resources of the National Research Council.

The members of the technical committee selected to monitor this project and to review this report were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the project. The opinions and conclusions expressed or implied are those of the research agency that performed the research, and, while they have been accepted as appropriate by the technical committee, they are not necessarily those of the Transportation Research Board, the National Research Council, the American Association of State Highway and Transportation Officials, or the Federal Highway Administration of the U.S. Department of Transportation.

Each report is reviewed and accepted for publication by the technical committee according to procedures established and monitored by the Transportation Research Board Executive Committee and the Governing Board of the National Research Council.

The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and of advising the Federal Government. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine. The National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively, under the charter of the National Academy of Sciences.

The Transportation Research Board evolved in 1974 from the Highway Research Board, which was established in 1920. The TRB incorporates all former HRB activities and also performs additional functions under a broader scope involving all modes of transportation and the interactions of transportation with society.

Published reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from:

Transportation Research Board
National Research Council
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

and can be ordered through the Internet at:

http://www.nationalacademies.org/trb/bookstore

Printed in the United States of America

PREFACE

A vast storehouse of information exists on nearly every subject of concern to highway administrators and engineers. Much of this information has resulted from both research and the successful application of solutions to the problems faced by practitioners in their daily work. Because previously there has been no systematic means for compiling such useful information and making it available to the entire community, the American Association of State Highway and Transportation Officials has, through the mechanism of the National Cooperative Highway Research Program, authorized the Transportation Research Board to undertake a continuing project to search out and synthesize useful knowledge from all available sources and to prepare documented reports on current practices in the subject areas of concern.

This synthesis series reports on various practices, making specific recommendations where appropriate but without the detailed directions usually found in handbooks or design manuals. Nonetheless, these documents can serve similar purposes, for each is a compendium of the best knowledge available on those measures found to be the most successful in resolving specific problems. The extent to which these reports are useful will be tempered by the user's knowledge and experience in the particular problem area.

FOREWORD

By Staff
Transportation
Research Board

This synthesis report will be of interest to both administrators and researchers involved in the measurement of performance for research, development, and technology (RD&T) programs in state departments of transportation (DOTs), federal transportation and other agencies, the private sector, and academic organizations. It addresses the use of performance measures for evaluating the effectiveness and impact on transportation RD&T programs. The synthesis is based on a review of the literature, a questionnaire distributed to state DOTs (research programs, upper management, and operating units), and queries to select private sector and academic organizations.

Administrators, engineers, and researchers are continually faced with highway problems on which much information exists, either in the form of reports or in terms of undocumented experience and practice. Unfortunately, this information often is scattered and unevaluated and, as a consequence, in seeking solutions, full information on what has been learned about a problem frequently is not assembled. Costly research findings may go unused, valuable experience may be overlooked, and full consideration may not be given to available practices for solving or alleviating the problem. In an effort to correct this situation, a continuing NCHRP project has the objective of reporting on common highway problems and synthesizing available information. The synthesis reports from this endeavor constitute an NCHRP publication series in which various forms of relevant information are assembled into single, concise documents pertaining to specific highway problems or sets of closely related problems.

This report of the Transportation Research Board provides information on performance measures as they relate to RD&T programs. It addresses the general issue of measuring performance, and it details various activities carried out within state DOTs, federal and other agencies, the private sector, and academic organizations. It serves as a tightly

focused companion document to *NCHRP Synthesis 238*, and is an outgrowth of the activity involved with *NCHRP Synthesis 280*.

To develop this synthesis in a comprehensive manner and to ensure inclusion of significant knowledge, the available information was assembled from numerous sources, including a large number of state highway and transportation departments. A topic panel of experts in the subject area was established to guide the author's research in organizing and evaluating the collected data, and to review the final synthesis report.

This synthesis is an immediately useful document that records the practices that were acceptable within the limitations of the knowledge available at the time of its preparation. As the processes of advancement continue, new knowledge can be expected to be added to that now at hand.

CONTENTS

ACKNOWLEDGMENTS

Scott A. Sabol, Vermont Technical College, Randolph Center, Vermont, was responsible for collection of the data and preparation of the report.

Valuable assistance in the preparation of this synthesis was provided by the Topic Panel, consisting of William G. Agnew, General Motors (retired), Washington, Michigan; Carolyn D. Goodman, Technology Transfer Director, Virginia Transportation Technology Transfer Center, Virginia Transportation Research Council; Mary Jo Hamman, Indiana Department of Transportation; Robert C. Kelly, Research Program Manager, Federal Highway Administration; Mark Norman, Director, Technical Activities, Transportation Research Board; Morris Oliver, Transportation Specialist, Federal Highway Administration; Martin D. Pietz, Director of Transportation Research, Washington State Department of Transportation; Dr. Deniz Sandhu, Manager, Technology Transfer Section, Transportation Research and Development Bureau, New York State Department of Transportation; and Larry A. Scofield, Researcher–Transportation Research, ATRC/Arizona Department of Transportation.

This study was managed by Stephen F. Maher, P.E., Manager, Synthesis Studies who worked with the consultant, the Topic Panel, and the Project 20-5 Committee in the development and review of the report. Assistance in project scope development was provided by Donna L. Vlasak, Senior Program Officer. Don Tippman was responsible for editing and production. Cheryl Keith assisted in meeting logistics and distribution of the questionnaire and draft reports.

Crawford F. Jencks, Manager, National Cooperative Highway Research Program, assisted the NCHRP 20-5 Committee and the Synthesis staff.

Information on current practice was provided by many highway and transportation agencies. Their cooperation and assistance are appreciated.

PERFORMANCE MEASURES FOR RESEARCH, DEVELOPMENT, AND TECHNOLOGY PROGRAMS

SUMMARY

Although performance measures of one sort or another have been around for generations, in the last decade they have become more popular, especially among public agencies. The genesis for this popularity was the passage of the Government Performance and Results Act of 1993, which directed federal agencies to develop formal mechanisms for measuring and reporting performance. This popularity spread across the nation to state and local agencies, including state departments of transportation (DOTs).

In the last decade, many studies related to performance measures have been undertaken, and performance measures have been implemented in many of the production units of state DOTs. However, performance measures for research, development, and technology (RD&T) programs have been slower to be implemented. This is for several reasons, not the least of which is the difficulty of measuring the quality or value of research activities.

Agencies that have instituted or are investigating performance measures for RD&T are typically focused on the following three concerns of the research program:

- Process management,
- Program quality, and
- Program value.

Many performance measures for process management have been in place for many years. Far fewer are in place for program quality or value. For program value, benefit-cost ratios and the use of anecdotal evidence of success are the most common.

Examples of performance measures used by state DOTs for project selection include cost-benefit analyses, qualitative alignment of proposed projects with organizational strategic goals, and peer assessments. For project management, the typical performance measure is related to benchmarking actual progress against the prescribed budget and schedule, often making use of a database and updating records as needed (where projects that are significantly deviating from anticipated progress can be easily recognized). For post-project implementation, a typical performance measure is of a binary nature—a "check-off" approach that asks if units within the DOT were using the results of the research. Program benefits are often assessed using cost-benefit analysis techniques (with great variation among states regarding the specifics of implementing the technique) or through anecdotal evidence (e.g., reports from operating units within the DOT) of payoff from the research. Finally, there was little in the way of performance measurements regarding staff productivity related to the RD&T effort—most agencies that had a method in place indicated that it was the general performance review that applied to all DOT staff.

Investigation has indicated that for the often intangible and intractable benefits of research, peer-assessment processes are the most effective means of obtaining measures of program performance. This ideal has instant credibility among state DOTs and the U.S. DOT, who have embraced the concepts of "peer exchanges," which rely on interactions of external parties (often colleagues in other DOTS) with their programs. These peer exchanges, meant to foster open dialogue among colleagues in the RD&T realm across agency boundaries, and not intended to be formal assessments, but rather information exchanges, have been promulgated through FHWA activities and through continued involvement in the peer exchange programs conducted by state DOTs in cooperation with the FHWA. In addition, the National Science Foundation continues to allocate significant portions of its funding based on peer-assessment processes. Peer assessment has been given further credence based on recent work of the Committee on Science, Education, and Public Policy of the National Research Council.

The private and academic sectors also struggle with performance measurement, and a viable solution for state DOTs does not reside with them. In many ways, private sector practices are similar to those of the few public RD&T units with comprehensive measures in place. Many of the concerns within the private sector regarding costs, savings, and improvements to operations are shared by the public sector.

In contrast, the academic sector historically has been focused on matters far different from those of the sponsors of research. Although their performance measures are limited in portability to the RD&T unit, an understanding of these measures is essential—after all, the academic community provides the lion's share of the work force when it comes to the conduct of research and education/training efforts for many state DOTs (a significant contribution to federal transportation activities is also provided).

It is clear that tying some performance measures to the strategic goals of an agency is essential. This linkage ensures the relevance of the research unit to the desires of upper management. However, the research unit must have self-contained performance measures as well, because the strategic goals of transportation agencies are seen by many to change as frequently as do administrations; however, RD&T programs require longitudinal performance data that can be compared over much longer periods of time.

It is also clear that much remains to be done before adequate tools are available to satisfy the needs of research managers related to performance measurement. Specifically, guidance is needed on cost-benefit approaches to program-level evaluation, and tools are needed to assess research outcomes that do not lend themselves to quantitative measure (such as anecdotal stories of success, institutional knowledge capacities, and similar factors). A specific need lies in the area of technology transfer, such that the extent of the improvement of workshop or training session attendees' skills or abilities can be more rationally determined.

CHAPTER ONE

INTRODUCTION

Performance measures became one of the catch phrases of the 1990s as state agencies came under increasing scrutiny with respect to their budgets, operations, and business operations. Many state agencies attempted to implement performance measures to help evaluate the quality and quantity of work produced by its operating units. State departments of transportation (DOTs) were no different, often setting benchmarks for production (e.g., designs or constructed projects completed). However, although there were some states that developed or already possessed performance measures for their research, development, and technology (RD&T) programs, many did not. Indeed, to date, approximately one-half of the state DOTs surveyed as part of this synthesis effort did not have formal measures in place. Most, however, acknowledged the need for such measures in the short or long term.

This synthesis discusses performance measures as they relate to RD&T programs. It addresses the general issue of measuring performance, and it details the various activities carried out within state DOTs, federal and other agencies, the private sector, and academia related to evaluating the effectiveness of RD&T efforts. The synthesis is based on a review of the literature, a questionnaire to state DOTs (research programs, upper management, and operating units), and queries to selected private sector and academic organizations.

WHAT ARE PERFORMANCE MEASURES?

Performance *measures* are assessment data or techniques that strongly, directly, or quantitatively reflect the degree to which results meet the needs and expectations of the customer (*1*). These measures are often compared to goals or benchmark levels, such that remedial actions can be activated when benchmarks are not met. Performance *indicators*, on the other hand, are data or techniques that suggest general alignment of results with customer goals. They can typically be direct or surrogate measures for actual performance characteristics of interest, and they often are useful in identifying trends in overall performance, if not an actual comparison to a desired goal. In simplest terms, performance measures are a stricter, quantitative benchmark of results compared to performance indicators. With care, performance measures can be aggregated from local to state to regional to national levels. Some performance measures may even be presented in a way that allows the state DOT to be compared with other agencies within state government (e.g., if some measure based on dollars is used).

A contrast of the terms *measure* and *indicator* is available if one considers a ubiquitous item tracked by many DOTs: anecdotal stories about RD&T "successes." The existence of unsolicited anecdotal stories from other DOT units, reporting how RD&T helped them, might be considered a "performance indicator." However, a "performance measure" might be the actual number of such stories recorded per year, with a benchmark or goal of two. Anecdotal evidence provides an intuitive and subjective connection between activities and benefits in a way that is not easy to capture quantitatively.

Performance measures are used in many facets of life. For example, an automobile manufacturer may measure the quality and execution of its function, car making, using performance measures such as number of recalls or defects noted after delivery to dealers or number of cars produced per day. Performance measures in state DOTs have gained increased use in the last decade, addressing operational issues such as the number of roadway designs completed or the extent of shipping activity in ports. *NCHRP Synthesis of Highway Practice 238: Performance Measurement in State Departments of Transportation*," published in 1997, provides a comprehensive discussion of the terminology and state of the practice regarding performance measures for DOT operating functions (*2*). However, the document contains little information directly related to performance measures for RD&T programs in state agencies.

The term performance measures has been around for some time, but its popularity has increased with the passage in 1993 of the Government Performance and Results Act (GPRA). Although there are various traditional sources of definitions of additional terms related to performance measurement, no standard and commonly accepted definitions currently exist. Since 1995, the U.S. Office of Management and Budget has provided online a "Primer on Performance Measurement" (*3*). The primer provides definitions that are intended for use in federal programs in adherence to GPRA. The following terms are included, although many state DOTs do not use the same terminology:

- Outcome measure—An assessment of the results of a program compared to its intended purpose (e.g., percentage of taxes collected in a given year by a revenue agency).
- Output measure—A tabulation, calculation, or recording of activity or effort that can be expressed in a

quantitative or qualitative manner (e.g., timeliness of payments to consultants on design projects).
- Impact measure—Measures of the direct or indirect effects or consequences resulting from achieving program goals (e.g., decrease in design errors resulting from training of state DOT engineers).
- Input measure—degree or extent of resources or other items that an agency or manager has available to carry out the program or activity (e.g., number of employees available to do planning studies).

Perhaps because of the differing understandings of performance measures, New York established some definitions and usages under *Operational Goal 94-8*, which it published in a Technical Services Division document, "Review of Performance Measures" (*4*). This document, which does not apply only to the research unit, defines a performance measure as "a management tool for objectively gauging the work accomplished towards a goal." Furthermore, the document establishes that managers use the measures "as a tool for

- Effectively utilizing their resources,
- Measuring past work,
- Rating work against standards,
- Comparing similar groups, and
- Reporting work output and performance."

In addition, it notes that such measures provide managers with facts "that are used to

- Determine if goals are being met,
- Defend present resources,
- Justify additional resources,
- Measure efficiency, and
- Improve performance."

WHY ARE PERFORMANCE MEASURES IMPORTANT TO RD&T PROGRAMS?

Performance measures have become one component of public and internal perception of accountability—an issue that has recently grown in prominence within state DOTs, as noted in *NCHRP Synthesis of Highway Practice 280: Seven Keys to Building a Robust Research Program* (*5*). RD&T programs are often perceived as unusual entities within an organizational structure, with special concerns related to their accountability. Sometimes the programs are perceived as being treated "with kid gloves"; that is, not being held to the same standards as the rest of the agency.

The issue of accountability goes beyond mere perception. Although considered essential to the vitality of an organization, RD&T programs are usually one or more steps removed from the "front lines" of organizational activity. This is especially true in transportation RD&T programs. Frequently, the results of RD&T activity are used by others to improve operations. The unit that uses the results will often have readily measurable performance outcomes, although the role and extent of RD&T's impact on those outcomes is harder to delineate. For example, consider the following performance measures discussed in *NCHRP Report 446: A Guidebook for Performance-Based Planning* (*6*):

- Accessibility
- Mobility
- Economic development
- Quality of life
- Environmental and resource conservation
- Safety
- Operational efficiency
- System preservation.

Many of these measures are directly applicable to the activities of operating units in a state DOT structure. For example, accessibility and mobility increases may be linked to planning activities and goals, operational efficiency to a maintenance or construction unit, and safety to a design unit. RD&T activities may have contributed to some or all of these areas; however, the linkage is not direct and the actual improvement based on the RD&T activity alone (i.e., setting aside any contributions solely from the operating unit) is very difficult to ascertain.

Increasingly, chief executive officers (CEOs) of state agencies are being hired from the business sector, frequently from organizations that had no research and development (R&D) unit. Accordingly, they are sometimes perceived as being without an in-depth appreciation for the special role of RD&T, and they may not have an appreciation, or perhaps a tolerance, for a unit that operates under special conditions and understanding.

In addition, based on discussions with DOT personnel during the gathering of information for this synthesis, it is clear that RD&T programs have not been spared from the attrition rates of other units in state DOTs. With such attrition (through early retirement and similar programs), the institutional memory of past performance of the programs is diminishing. Although unspecified goals and measuring sticks may reside in the head of the research administrator, there is a risk that such corporate wisdom may disappear with the individual. Accordingly, there is an identified need to have performance measures be a process-oriented activity for the continuous improvement of the program, and not a system that is based solely on a particular person at the top of the RD&T organizational chart.

For these and other reasons, RD&T units may choose to embrace performance measures to help increase their chances for improving and surviving in the current transportation climate.

TERMS USED IN THIS SYNTHESIS

As noted previously, there are no commonly accepted definitions for terms used related to performance measurements. However, many terms are ever present, and some of these are subsequently described with their usage in this document. Some of the distinctions may not be the same as used in all agencies; however, the subtlety of some distinctions will not detract from the primary message of this synthesis.

Research (Basic and Applied)

Perhaps one of the most important questions that is or should be asked about RD&T programs is "What is research?" Surprisingly, there are many differing opinions and interpretations related to this term. Most state agencies and others accept variations of the traditional dictionary definition, one of which states that research is "studious inquiry, usually critical and exhaustive investigation or experimentation having for its aim the revision of accepted conclusions, in the light of the newly discovered facts" (*7*). Many CEOs and other administrators do not hold to a strict definition of research; instead, research is defined as simply something that requires more than standard operational time to conduct or implement, but not necessarily the revision of understanding and not necessarily performed under any standard operating parameters such as the "scientific method." Most persons involved in the research, be they managers, academics, or others, tend to require a strict definition, with an informal use of the term being considered some sort of a "special consulting study" or "special planning study" but not necessarily "research."

Under the rubric of research, there are two commonly accepted although not mutually exclusive branches: basic and applied. These distinctions are more commonly agreed upon than the actual definition of research itself. As stated in a recent study from the National Academy of Sciences (*8*), "basic research is often thought of as an unfettered exploration of nature whose only required output is new knowledge and whose outcomes are unknowable in advance. Applied research might be described as an activity whose outputs are also new knowledge, but knowledge whose nature and use are explicitly needed to achieve a specific useful outcome."

An example of basic research might be the pursuit of mathematical formulations for complex or other analytical functions, where the solutions may never be found. The pursuit of a mathematical proof that has stumped scholars for years (e.g., Fermat's Last Theorem) can be categorized as basic research—it may nurture entire branches of mathematics through the creative meanderings of mathematicians over the centuries. Applied research might include the creation of numerical techniques (perhaps through computer code) that provide approximate solutions to mathematical formulations known to describe physical processes.

State DOTs and the private sector tend to focus almost exclusively on applied research, whereas national initiatives from government or foundations will frequently sponsor both applied and basic research. Some states do perform or sponsor basic research, sometimes with a hope for eventual relevance to their activities, and sometimes because it provides enhanced technical expertise and technical networking abilities to their staffs or pool of researcher talent.

Development

Development is frequently seen as the process required to take an idea or knowledge and to manifest it as something physical or real (with some deference for the world of software development with respect to the term "physical") and useful. Development is different from production in that the path from idea to manifestation may not be direct or prescribed, and it may diverge significantly from an originally anticipated final product. Nonetheless, successful development inevitably provides a product that meets originally stated goals or objectives.

Many state DOTs are actively involved in development—for example, the creation of new mixture formulations for pavements that achieve prescribed performance goals.

Technology and Technology Transfer

In general, technology relates to concepts, practices, and developments in industrial science or the industrial arts (*7*). For this synthesis, the focus will be on technology transfer, or the conveyance of such understanding to a less informed individual or group. Many state DOTs have as a major need the training of new and existing personnel. The technology transfer unit, whether it is a "T^2 center" or local technical assistance program (LTAP), often meets this need. Frequently, the technology transfer is intended to bring employees up to speed on new ways of performing job functions, where these new ways have resulted from research and/or development.

Science and Engineering

Science generally is concerned with knowledge related to observation and classification of facts and/or the establishment of verifiable general laws. For example, the world of mathematical science describes, in part, interactions of

functions based on sets of commonly accepted axioms. Engineering, on the other hand, relates to the use of scientific concepts in order to understand or to mold pieces of the physical world to meet prescribed needs, with the process adhering to economic and social constraints in addition to the physical laws of the universe. State DOTs frequently sponsor studies in both science and engineering.

Qualitative Versus Quantitative Measures

Qualitative measures are those that can be described in a meaningful way, yet escape any specific numerical or hierarchical relativity. Frequently, qualitative measures are subjective, and a specific measure (e.g., this riding surface is good) may have a different meaning depending on who is providing the interpretation. Quantitative measures, on the other hand, are those that can be assigned some numerical or hierarchical relativity. For example, pavement condition may be, in part, assigned a quantitative measure of its ride quality based on the International Roughness Index. There are commonly accepted methods for computing this index (e.g., ASTM E1926-98, "Standard Practice for Computing International Roughness Index of Roads from Longitudinal Profile Measurements") and the numerical (quantitative) result is generally perceived to be objective and understandable among various members of the worldwide pavement engineering community. However, these same members might not agree on what a "good" road is.

In general, state DOTs are involved in activities that lend themselves to the use of these terms and the performance measurements that are developed should be consistent with the activities being measured or objectives being sought.

WHAT IS THE GENESIS OF PERFORMANCE MEASURES?

The first performance measure probably dates back to prehistoric times, where successful performance against the environment resulted in survival. The recent attention to performance measures in government agencies can be traced back to the 1980s and early 1990s, culminating with the passage of GPRA. The reasons for the attention to performance are many, and they can be tied to at least the following topics.

Possible Outcomes from RD&T Efforts

RD&T efforts may provide many outcomes that are considered useful to both the DOT and the traveling public. Among these, the following are often cited as important "results" from research activities:

- Cost and resource (energy) savings,
- Innovations,
- Improved safety, and
- Improved customer satisfaction.

Performance measures allow for a quantification of these results, a comparison of current activity to past or planned activity, and a basis for reporting successes to customers and others.

There are other ways of thinking about outcomes from R&D; for example, the following are perceived to be four impacts of R&D (J. McEntire, U.S.DOT, personal communication, April 27, 2001):

- Creation of public goods,
- Front end activity for capital acquisitions (R&D can set the stage for good long-term procurements),
- Providing a proper foundation for regulatory requirements and policy, and
- Ensuring knowledge and institution capacity ("human capital") to perform.

Again, performance measures provide for the quantification of these impacts and help to provide RD&T systems that continuously improve.

Evolution Within the Transportation Community

Because of the creative nature of research and the generation of new knowledge, methods, and technologies, RD&T efforts serve to spur the evolution of the transportation community. New knowledge or technologies that are useful become integrated into practice and allow for the overall improvement of the transportation system or those involved in the operation of the system. Knowledge that critically analyzes existing technologies or approaches may result in the early abandonment of such practices, allowing for reallocation of resources among more promising or profitable activities within an organization.

Regulatory and Administrative Basis

Regulations and administrative policies typically flow from the will of the people or from a personal belief. It can be inferred that the passage of GPRA was a congressional response to perceptions of the will of the people; that is, that there should be specified directives on the accountability of federal programs. Such codified requirements were implemented through regulatory channels, and many state or local agencies adopted such policies (either based on a belief that it was a good idea or because the local legislature had enacted similar requirements).

These topics (possible outcomes, evolution of the system, and administrative/regulatory basis) all are related to a perceived benefit from a structured reporting of how pro-

grams are meeting objectives and goals. All three, independently and jointly, helped to foster a climate for the advent of performance measures for RD&T programs.

Background on the Roots of the Current Study

This synthesis arose from discussions and efforts within TRB committees that deal with the issues of management of RD&T programs, primarily Committee A5001 on Conduct of Research. This committee is comprised of individuals who perform research, manage research, and use research results. There was committee cognizance of the "onslaught" of performance measures in many operating units within state DOTs, and concurrent recognition that many research programs were ill-prepared to implement adequate measures, while other programs were well along in such implementation. Accordingly, Committee A5001 recommended preparation of a document that could provide guidance on performance measures for RD&T programs, providing background information on issues and tools.

SCOPE AND AUDIENCE OF THE REPORT

Public sector transportation RD&T programs recognize the value of employing performance measures to track the effectiveness and impact of these research programs. This synthesis addresses uses of performance measures for evaluating the effectiveness and impact on transportation RD&T programs. It is based on a literature review and a survey of practice, and attempts to address issues such as

- The various kinds of performance measurements being used,
- How they are developed,
- How effective the measures are in evaluating individual RD&T transportation programs,
- The commonality of measures across various programs, and
- Best practices.

It is important to note that this document is intended to assess what is currently available regarding performance measures—it is not the intent of this synthesis to develop or recommend specific performance measures to fill gaps in existing practice. However, the document does contain recommendations related to work needed to fill in such gaps in the future.

The synthesis is targeted to managers of RD&T programs, as well as others who are interested in assessing the performance of such programs. It is written to provide a background on performance measurements, some specifics about actual measures in use, and to indicate some recommendations for additional work in the area. The synthesis serves as a tightly focused companion document to *NCHRP Synthesis 238* (*2*).

REPORT ORGANIZATION

This synthesis is divided into several chapters, some of which may be read independently of others.

- Chapter 1 introduces the topic of performance measures, identifying what they are and some of the common terminology.

- Chapter 2 provides some specific history on performance measures for RD&T programs, with clear indications of the lack of readily available information, whereas chapter 3 provides some background information related to issues surrounding performance measurement. For some readers, this latter chapter may serve as a primer on performance measurement concepts.

- Chapters 4–7 address performance measures for RD&T programs in state DOTs, federal and quasi-federal agencies, the private sector, and academia. These chapters provide the basis for comparison and contrast among these four realms, and provide indications of the linkage and disconnects that exist.

- Chapter 8 provides a summary of commonalities among the groups noted in the preceding four chapters, whereas chapter 9 presents conclusions based on the information reviewed for the synthesis and recommendations for future activities.

- Finally, references and appendixes are provided that indicate the sources of information, present questionnaires and tabulated responses, and illustrate other information noted in the text.

CHAPTER TWO

HISTORY AND PURPOSE OF PERFORMANCE MEASURES FOR RD&T PROGRAMS

As mentioned previously, performance measures have been used in some form for a long time. However, it has been in the last 10 to 15 years that the topic has gained substantial notoriety. Discussed here are some of the historical aspects of performance measures in general, with special attention paid to those of RD&T programs, and some discussions of their purpose. For a more complete background on performance measures for state DOTs, information is available in several recently published sources (*2, 5, 7–9*). Much of the information related to performance measures in general is also relevant to RD&T programs in particular.

PERFORMANCE MEASURES AS A MEANS FOR IMPROVING AND JUSTIFYING PROGRAMS

Success of an RD&T program depends substantially on perceived value to the agency as a whole and on accountability. Simply put, a program that is vital and robust will typically be perceived as an asset, and one that is languishing may be perceived as worthless, or even worse, a burden to the rest of the agency. The question becomes, How does a research manager know if his or her program is robust or languishing? Given the challenges of measuring the payoff from research and the limited resources available to many programs, the answer often lies in the perceptions of the research manager. A program that can obtain objective evidence to indicate how well it is performing may be well ahead of the norm and can institute remedial actions if problems exist. Accordingly, a system of measures, and goals or benchmarks for those measures, becomes clearly desirable for providing sustainability and long-term improvement to the program.

In addition, operating units throughout many state agencies are continuously being asked to document their value to the agency. For many units, such documentation is handy and well understood by upper management. For example, a CEO can appreciate when told that 90 percent of the design projects targeted for bidding have been completed on schedule. However, the same CEO may not understand the ramifications, good or bad, to the agency's overall mission when the research manager reports that four reports have been published and one set of standard specifications has been revised based on the results of a recent project. This is not peculiar to CEOs—many researchers and research managers grapple with the same issue!

At the national level, entire technology transfer program strategies have been investigated, with the purpose of providing overall improvement at a program- or agency-wide level. Such issues are discussed in TRB *Special Report 256: Managing Technology Transfer: A Strategy for the Federal Highway Administration* (*10*).

In academia, R&D programs also look for constant improvement and justification. Frequently, program improvements result in increased funding from sponsors, and funding level is one performance measure used at research universities. Often, academic research can be in an area whose importance may not be known to the general research community, and an associated lack of appreciation for that research area can occur. However, such programs historically have been able to point to sustained and often high levels of funding that "prove" their worth to the unknowing outsider.

PERFORMANCE MEASURES IN OTHER ASPECTS OF DOTS AND AGENCIES

The history and purpose of performance measurements in DOT units other than RD&T has been treated comprehensively elsewhere (*2*) and is not repeated here. However, there is a family of other (non-DOT) agencies that have historically struggled with the issue of performance assessment—departments of education (state, local, or national). Considering that educational efforts in many ways mimic R&D efforts and are very closely aligned with technology transfer concerns in DOTs, it is clear that the historical path taken in the educational arena should be considered. Several reports and other documentation are available on the subject, providing background on past activities and on the relevant issues related to educational assessment (*11–13*). Most of these reports address the issue of the difficulty in quantifying the output from the educational process and the difficulty in evaluating such a large-scale operation. (This can be viewed as "On a one-to-ten scale, how does one gauge how well educated a student has become?" or "How can one system be applicable to millions of students in millions of different learning situations?") Of particular relevance are those performance measures in the "educational realm" that concern libraries, because the library function of an RD&T program (providing a storehouse of documentation, providing assistance with searches for information, etc.) is often one of the measures used by customers of the research program.

REGULATORY AND ADMINISTRATIVE BASIS

The Government Performance and Results Act of 1993, which stemmed from a tide of local and national discussions, both formal and informal, on the need for quantitative monitoring of activities to help provide both credibility and budgeting assistance, is probably the most well-known legal basis for today's work at national and state levels. GPRA presents requirements related to strategic planning and performance-based budgeting, and reporting requirements for performance.

IMPACTS OF PERFORMANCE MEASURES

Performance measures in RD&T programs can have several impacts, some of which are described in chapter 4 in a discussion of state practice. Impacts come under two broad categories: impacts on the *programs* and impacts on the *transportation system*. The existence of performance measures can affect programs in a positive or negative manner, especially as relates to the key program attributes identified in *NCHRP Synthesis 280* (*5*): a foundation of trust, bold marketing, ties to economics, unabashed deal making, accountability, inclusion of policy research, and staff empowerment. Having defensible performance measures, coupled with realistic goals, can provide, for example, instant credibility in the program and unquestioned accountability. Similarly, a robust RD&T program usually has, as an ultimate goal, the desire to have a positive and noticeable impact on the transportation system at large. When performance measures are in place that increase the likelihood, on average, of a specific project being completed on time, on budget, and providing the type of result that the customer requires, then the likelihood also increases, on average, that the transportation system will be enhanced by the research results. It is important to note that these two types of impacts are very different and may even be at odds with each other. Attempts to force on-time performance of research at the expense of quality may result in one performance measure that indicates program success, whereas another may indicate little or no improvement to the overall transportation system. Accordingly, performance measures must be balanced between these sometimes competing areas.

CASE STUDY—A SAMPLE PERFORMANCE MEASURE: MARYLAND STATE HIGHWAY ADMINISTRATION

Because performance measures can be relatively new to many state agencies, it may be useful to consider a specific performance measure in use at one state DOT. This helps to set the stage for the discussion of issues surrounding the development and implementation of performance measures that are described in the following chapter.

In August 2000, the Maryland State Highway Administration (SHA) developed performance measures for its RD&T programs. Several relate to key performance areas (KPAs) in Maryland SHA's business plan and include:

- Mobility
- Highway safety
- System preservation
- Economic development
- Community enhancement
- Environmental responsibility
- Customer service
- Resource management.

The SHA identified the following performance measures to track how the research program is meeting benchmarks and assisting with the overall mission of the agency.

- Percentage of research projects that can be applied to one of the KPAs noted above,
- Percentage of research projects for which benefits can be documented (a large percentage of projects with undocumentable benefits may be deemed undesirable for the program), and
- Quantify how research pays off.

Several things are noteworthy about the choice of these particular performance measures. First, they have all been tied to the strategic goals of the overall agency (through their relation to the KPAs). This helps to ensure that there is upper management support for and an appreciation of the performance measures. Second, the information being tracked is quantitative. Some of it may be objective (e.g., the first measure), whereas some may be somewhat subjective (e.g., the last). However, in all cases, an attempt is made to provide quantitative information that can be compared on a year-to-year basis or to make future adjustments. Finally, it is worth noting that these measures are not necessarily revolutionary! Indeed, this type of information may have been tracked, in one form or another, for years. However, it is clear that the SHA RD&T program has not chosen to pursue all of the possible measures related to the KPAs. Thus, they have presumably searched out the most significant matters and come up with a manageable number of items to track.

CHAPTER THREE

ISSUES

When embarking on an effort to develop or improve performance measures within an RD&T program, several issues must be considered. Although many may regard the following items as obvious, a clear understanding of them and their relationships to research programs is an essential step on the path toward successful implementation of performance measures.

DEFINING PERFORMANCE

Performance is defined as the execution of required functions. However, for RD&T programs, there are real questions related to the quality and extent of the execution, and what the required functions are or should be. Hence, there is an implicit reference to the manner in which actions are carried out that complicates matters. Many activities and processes, expectations, and perceptions as to outcomes exist. Accordingly, one of the first things that a research program must grapple with is its own definition of performance.

In general, it is best if performance is strongly related to the strategic goals of the transportation agency. Thus, it may be premature for a program to establish performance measures if the agency has no overall strategic plan in place. However, definitions of performance should not be tied solely to strategic goals. Why is this so? Consider the number of different strategic directions that agencies may take with each new administration or CEO. If performance were tied solely to current strategic goals, and then the goals change, it would be difficult to monitor performance over the long term. Accordingly, performance in areas aligned with the strategic plan of the agency, as well as research programs, should also be studied in areas that are considered to be stable through changes in upper management, to ensure that the areas of stability become part of the strategic goals of the agency.

WHAT SHOULD BE MEASURED?

Research units are typically concerned with at least five areas of activity.

- Project selection
- Project management
- Post-project implementation
- Staff productivity
- Program benefits.

Typically, measures should focus on these items, because success in these activities tends to result in overall program success.

It may sound simplistic, but it is essential that programs only attempt to measure the measurable. If an agency has no clear understanding of what constitutes the quality of a project or a program, then there is little sense in measuring for it. Similarly, agencies should only measure a manageable number of the most important issues. Too frequently, managers or researchers are tempted to collect all the data possible and then sort out the important items later. Research managers should remember that any activity, be it for project management, program management and evaluation, or otherwise, requires staff resources. Any efforts to collect data on performance will take away time available for the day-to-day operation of the RD&T unit. Accordingly, care should be exercised in determining what or how much data are to be collected and analyzed.

Data to be measured should be data that are known or strongly suspected to have a significant impact on performance. Many research managers are familiar with multivariate functions and performing sensitivity analyses to determine the correlation of the outcome to each of the input variables. Performance measures operate in much the same way—it may be possible to track 10 items, but if 2 of those items account for 95 percent of the performance results, then one must question the usefulness of tracking the rest. The difficulty arises from determining or sensing which are the critical items to track.

By tracking few but substantial data, it is also simpler to initiate remedial actions when required and to obtain a good understanding of the ramifications. Again, if management policy is compared to multivariate mathematics, consider these two functions

$$F1(x_1, x_2) = 5x_1^{\,2} + 3x_2 + 0.4$$

$$F2(x_1, x_2, x_3, x_4, x_5) = 6x_1^{\,8} - 2.1x_2^{\,7} - 3x_3^{\,4} - 400x_4^{\,-3} - 0.05x_5$$

where x_i = measurable parameters for the RD&T unit.

The research manager can clearly get a better understanding of the behavior of his or her corporate unit from the $F1$ function than from the $F2$ function. If performance

can be significantly and accurately tracked with the simpler function, it may be more desirable, even if the more complex function provides a slightly more refined and precise result.

The previous discussion centers on measuring information that is useful based on a management science or logic approach. However, there is another area of information that may be worth measuring, depending on a particular agency's operating environment—measures based on *political realities and upper management desires*.

Results from the questionnaire conducted for this project, discussed in the next chapter, indicate that CEOs frequently measure the RD&T unit's performance based on measures other than those used by the unit itself. Sometimes it is not clear that the upper management measures are truly indicative of performance, but if the CEO needs certain information upon which to make judgments, that information probably should be collected to ensure the credibility and vitality of the program. It is possible that some research managers have never asked their upper management the simple question: "What is the one thing that my unit can tell you that will most help you to judge our performance?" That one thing may or may not match the research manager's beliefs as to what is important, but it is essential that strong consideration be given to collection and reporting of information deemed important by the CEO.

HOW CAN IT BE MEASURED?

Once a determination has been made regarding *what* to measure, the next question becomes "*How* will I measure it?" Considering the types of measures and methods of measurement helps to answer this question.

A comprehensive discussion on measuring performance that is useful and easy to understand is provided by Brown (*14*). This book describes the characteristics of an effective measurement system as including

- A low number of key measures,
- Measures that reflect both past performance and are able to address current and probable future performance,
- Measures must be reviewed and modified on a regular basis to reflect changes in the environment or in the agency's strategy, and
- That care should be taken in setting targets for each measure.

The book also discusses the flaws that are typical in measurement systems, including

- Too many measures are included,
- The focus is on solely short-term issues,
- The use of measures focusing on behavior instead of outcomes, and
- The use of measures that discourage teamwork (for example, measures that address individual goals and benchmarks that might not result in synergy or agency-wide improvement).

For types of measures, there are at least two groupings to consider.

- Direct measures versus indirect measures and
- Qualitative versus quantitative measures.

Direct measures are those that provide an explicit indication of the status of an item or datum. For example, if "percentage of on-time project completion" is an important performance measure, a direct measurement is possible by comparing the number of projects completed on schedule with the total number of projects. Indirect measures, on the other hand, provide a surrogate representation of the status of an item. Although direct measures are preferable, they are sometimes nonexistent or very expensive or difficult to determine, and indirect measures must suffice. An example of an indirect measurement in a research program may be the tracking of anecdotal compliments about specific projects from RD&T customers as a measurement of program success or quality. Compliments do not define success or quality, but it is fair to say that the existence of a preponderance of compliments suggests or implies success.

Qualitative and quantitative measures were previously defined in chapter 1. Generally, for management purposes, quantitative measures are preferred, because of their portability and understandability among various reviewers of information. However, there are cases for which quantitative measures are difficult or expensive to obtain, and qualitative measures provide reasonable guidance regarding a program's effectiveness or success.

Care must be taken to not mistake a qualitative measure for a quantitative measure simply because the qualitative measure has a number associated with it. For example, during the project selection process, many state DOTs rank the desirability of specific projects based on their likelihood of success. Although one state may rank the likelihood of the success of proposed projects as "low," "medium," and "high," another state may use a numerical system and rank the same projects on a scale of 1 to 3. The substitution of a number for a phrase does not make a measure quantitative. It is the subjectivity of a measure that makes it qualitative. If an agency could assign objectively determined rankings of likelihood of success, then the same 1, 2, or 3 above could be a quantitative measure.

Usually, quantitative measures can stand up to some use of statistical and mathematical analyses. This is much

harder and possibly inappropriate to do with qualitative measures, although newer analytical techniques, such as fuzzy set theory, are making it easier.

It is important to note that just because qualitative measures are sometimes not preferable, this does not mean that they should not be employed. If a particular indicator has a strong correlation to performance, and if that indicator can only be tracked using qualitative methods, then it should be part of the evaluation process.

There are also specific methods for measurement that are frequently employed by research agencies and these can generally be categorized as follows:

- Database methods
- Personal assessment
- Peer assessment
- User/client surveys.

Database methods are those methods that use tabulation and structured organization of numerical data. Such methods are commonly used in all facets of DOT operations, including in the research unit. Such methods are especially useful when the data are collected for general management reasons or as a matter of course. An example of a database method was described previously—tracking on-time performance. Most programs track the proposed schedules for projects against actual progress. Frequently, this is a quantitative measure that can be expressed in binary form (on schedule or over schedule) or as a percentage (e.g., completion at 95 percent of proposed, completion at 122 percent of proposed, etc.). Database methods are also frequently used for post-project implementation, although it may be in a simple binary format (e.g., "The results were incorporated into the standard specifications" or "The results were not used").

Personal assessment relates to the efforts of a single individual in evaluating an item and assigning a quantitative or qualitative measure to it. Although not all DOTs operate in this manner, several perform cost-benefit analyses of projects based on personal assessment. An individual, usually a research staff member, determines the overall costs of a specific project (this is often simplified to the contract cost for those agencies performing contract research) and estimates the benefits that result from the implementation of the project results or from the use of the knowledge obtained from the research. Although the staff member may get assistance from others in trying to estimate benefits, and although the staff member may use a prescribed methodology (such as that used by New York State), the final tally is typically a reflection of the work of a single person.

Personal assessment is a very common method to use for qualitative measures, because of its relative ease and usefulness. For example, in its periodic reports to its sponsors, the NCHRP provides a qualitative estimate of the probability of success of each project (low, fair, good, etc.). Although NCHRP staff base their decisions on common elements (past successes of the research contractor, perceived difficulty of the project, etc.), there is not a "level playing field" among staff for assigning a probability for success. However, the lack of equity among the ratings does not result in any significant problem, because the probability of success rating is qualitative information for the sponsors and is not an essential measure used regularly in any decision making for the program (although it could theoretically be used as a foundation for cutting off future sponsorship).

Peer assessment uses the input from multiple persons to come up with a final average or consensus measurement. Peer assessment is very useful when one is dealing with qualitative measures for which there may be wide variations in ratings from one individual to another. It benefits to a degree from the same concept as Gaussian distributions and the Law of Large Numbers—that is, in a large group of well-informed individuals, ratings on a particular item will tend toward a real mean, even with significant outliers. That is to say, in a group of 100 people, there will be some common ground among the group as to what constitutes a ranking of fair, good, or excellent. Although the perceptions may vary slightly among many in the group, and may vary significantly for a few, there exists a useful "herd mentality."

Peer assessment, sometimes referred to as expert review, is considered the most effective means of evaluating federally funded research programs, according to a 1999 report from the Committee on Science, Engineering, and Public Policy of the National Research Council (*15*). [This point was reinforced in the 2001 follow-up report (*8*)]. Peer assessment gained great notoriety and respect with the advent of the FHWA's "Peer Exchanges" of RD&T programs for state DOTs that receive State Planning and Research (SPR) funds. Under the Peer Exchange program, various issues affecting performance are considered and commented on by an objective peer group (comprised of staff from other state DOTs, the FHWA, academia, and elsewhere), and documented findings and recommendations are supplied to the RD&T unit. The objective is not a formal audit or assessment by the external parties, but instead an open dialogue and exchange of information. Nonetheless, the exchanges serve as a de facto assessment, albeit without some of the concerns that might accompany a formal audit process. State research managers were very sensitive to the need for an open exchange of information and ideas without the characteristics of an assessment. The peer assessments have been so successful that, frequently, the results of the Peer Exchange are immediately conveyed to upper management for consideration. At national meetings,

the usefulness of the Peer Exchange process for RD&T units has been stated many times. The national Local Technical Assistance Program (LTAP) leadership has recently embraced the Peer Exchange concept for use in LTAP and T^2 applications as well.

User/client surveys typically are formal processes in which a questionnaire is used to solicit feedback from those parties that work with the research unit. These forms are usually easy to administer and can provide good anecdotal information (they are good at providing qualitative information, but less so at quantitative). Such anecdotal information is often useful for sharing later with government officials or decision makers. User/client surveys often serve a secondary purpose as well—they are marketing and promotional tools for the RD&T unit.

COMMONALITIES AND DIFFERENCES AMONG MEASURES FOR RD&T

In general, research for this synthesis focused on R&D programs, with some attention paid to technology transfer functions in government agencies. Common among all three functions (RD&T transfer) is the notion that some of the results are simply very hard to measure in comparison to the functions performed by other units within a state DOT. Dollars (costs, benefits, or some ratio thereof) can be a useful, but not completely satisfactory, measuring stick. The number of projects completed, which can be a significant indicator of performance for, say, a construction unit in a DOT, in which project quality is relatively uniform, may have little reflection on the true contribution or success of the RD&T unit. Accordingly, all three functions must deal with a fuzzier world than their companion operating units.

The three functions also share an important but often underrecognized common outcome that must be measured—the development of the next generation of the transportation work force. Through contract research at universities, state DOTs are exposing students to specific and applied problems in the transportation world, and thus are influencing their potential career paths and their knowledge base about transportation systems. Similarly, through training programs, technology transfer efforts are improving and extending the understanding of existing transportation professionals.

The three functions also have significant differences, among which are federally mandated reporting requirements (possibly performance measures)—the accountability requirements for R&D under the federal SPR program differ from those of the LTAP programs. Because the functions must keep track of these performance requirements (see the previously discussed political realities and upper management desires), they, by necessity, must look at specific measures not shared by the other functions. Also, it must be recognized that in many states the LTAP function does not reside in the state DOT; accordingly, a mixture of issues for state RD&T programs that also provide technology transfer (often through implementation of research results) exists.

In addition, the outputs of the three functions are quite different. R&D may result in hard, tangible products (e.g., a new construction specification or a new bridge inspection instrument), whereas technology transfer generally influences only a person's thinking or manual skills. These are disparate outputs requiring different performance measures.

COST OF INFORMATION

If obtaining and analyzing information to gauge performance did not require reallocation of limited resources, then tracking "everything under the sun" might not be a horrendous idea. However, RD&T units are under the same pressures as other DOT units, especially in the last decade, to do more with less. Time used by a staff member to collect or analyze information for evaluating the performance of the unit is time not available for other, often frontline, activities such as project management, project selection, or post-project implementation. Because there is little point in evaluating performance if none of the frontline activities are being done, there is an occasional practice of investing little to no effort in program evaluation, apparently with the hope that the competent conduct of the required frontline work will serve in and of itself as an adequate performance measure.

The last item in the preceding paragraph is a bit of an exaggeration, but there is an expressed concern among state agencies about an unwillingness or inability to dispatch human resources for program performance evaluation efforts when the unit is already fully engaged in simply managing its work load. With these scenarios, it is very important to step back and take a studied approach to the benefits that may result from paying the costs to obtain information that might indicate performance levels or trends. It may be that an adequate, defensible assessment of performance, in terms understandable by the CEO, could result in additional resources. This is especially true if some functional relationship between resource availability and performance level can be established.

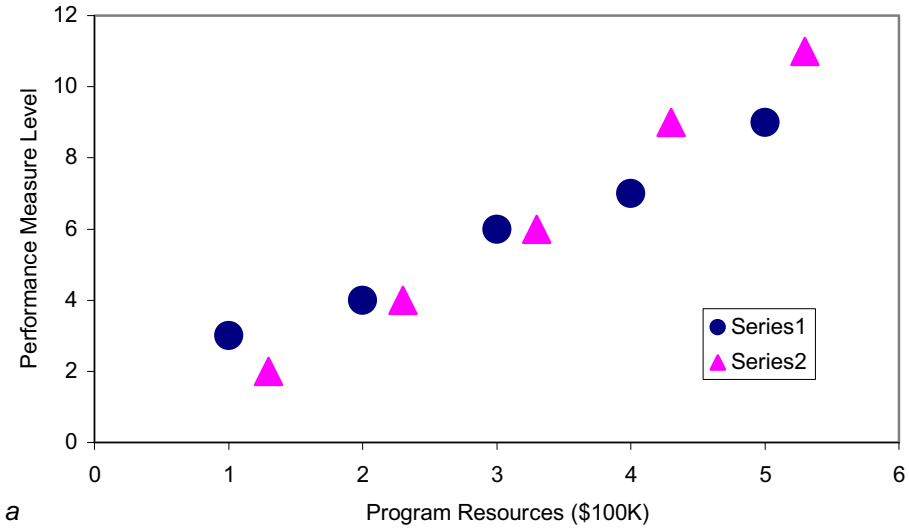
There are few resources available that provide good guidance on the cost of obtaining information. Some anecdotal evidence indicates that a single Peer Exchange can be conducted for approximately $3,000–$7,000 (depending on the number and location of participants). Information collected using database methods may be inexpensive if the

current system already tracks such information or if software tools for aggregating and analyzing the information are readily available. The costs of personal assessment methods depend greatly on the amount of effort required and the staff compensation rate on a case-by-case basis.

Given the variability in the costs of information, but the propensity for the cumulative costs to be high, great care should be given to using performance measures that are significant but not costly. Once a core group of performance measures is established, additional measures should be considered only after a determination that the costs of the collection of any additional required information will be outweighed by the usefulness of the information in enhancing the RD&T program.

RELIABILITY OF INFORMATION

The reliability of information or data and its effects on outputs or outcomes assessment has long been summed up by the phrase "garbage in, garbage out." An intricate analysis using qualitative measures from a personal assessment method may not be a wise combination given the inherent uncertainty in the inputs. Management actions based on analyses of performance measures and benchmarks should be taken with a clear understanding of potential statistical or other errors in the results. As an example, consider Figure 1, which shows some arbitrary data purely for the purposes of illustration. In Figure 1*a*, the average values of some performance measure are shown as Series 1 and Series 2, a function of overall



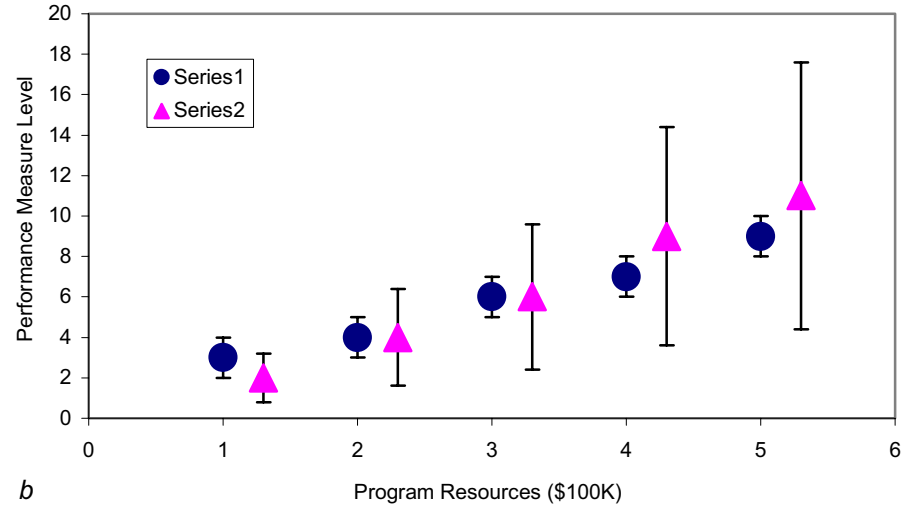


FIGURE 1 Demonstration of data uncertainty on performance measure interpretation.

program resources. Based on a quick scan of this graph, it would appear that Series 2 has a clearly superior performance level as additional resources are made available. However, in Figure 1*b*, error bars are shown with both series. Although Series 2 still trends higher, the error bars indicate that Series 2 may actually be a lower performer than Series 1. Although most RD&T programs do not perform rigorous statistical analyses of performance measures, it is extremely important to remember the impact of error (uncertainty), standard deviations, and other factors when attempting to graphically or otherwise portray information.

There is no "silver bullet" when it comes to matching an appropriate method of measurement to a performance measure, or when trying to decide which or how much data to collect. However, consideration of these matters may help state agencies to avoid some possible pitfalls.

CHAPTER FOUR

STATE DOT EXPERIENCES

Several state DOTs have been using performance measures, whether they have labeled them as such or not, for many years. Others are either only now instituting them or only just considering their implementation. This chapter focuses on summarizing the state of the practice among RD&T programs within the state DOTs. The information herein is based on personal communications, reports, presentations at conferences, and, primarily, on the results of a questionnaire that was distributed in 2000 to state DOTs. Approximately 60 percent of the states responded, with varying degrees of specificity in their responses. The questionnaire was in three parts: a comprehensive questionnaire to be completed by the RD&T unit, and two brief questionnaires, one to be completed by upper management, the other by customers (operating units) within the DOT. Not all parts were returned by all agencies. The questionnaires are provided in Appendix A, and a tabulation of the responses is provided in Appendix B. It should be noted that many states provided internal documents, strategic plans, or Peer Exchange reports to supplement their questionnaire responses, and this information was used to prepare the discussion that follows.

Approximately, half of the respondents indicated that they have no formal performance measures in place. Accordingly, these respondents strongly affect the perceived needs of the state DOTs, because they are starting from scratch. As can be seen from the questionnaire responses, a few programs have performance measures in place for virtually all functions. However, even these programs indicated needs for improvement for some or all of their performance assessments.

For the RD&T unit, the survey targeted the following primary concerns attributed to research programs:

- Project selection,
- Project management,
- Post-project implementation,
- Staff productivity, and
- Program benefits.

The upper management and customer questionnaires were more global and general, asking respondents about their perceptions of performance measures for the research unit, the perceived level of performance of the unit, and the basis for how they (management or customer) judged the RD&T program.

Research programs are typically interested in measuring performance with respect to three global ideals.

- Process management
- Program quality
- Program value.

Historically, RD&T units have focused on process management, primarily for two reasons: (1) the relative ease and low cost of focusing on this area, and (2) the match-up with typical business models in reviewing issues such as schedule and budget adherence. Moreover, it seems perfectly acceptable, and even desirable, to have performance measures related to process management if the public or your customer is measuring your unit on process-related items. In other words, process management may not have a significant impact on the overall contribution of the research unit's activities to the DOT's overall mission, but it may be considered unavoidable given some public perceptions of and concerns about governmental activities.

Process management measures are generally perceived as not sufficient to completely judge the worth of the research unit. There must be performance measures that address program quality (i.e., how "good" is the research, the staff, the researchers involved? and How meaningful are the projects that are selected?).

Overall, what many state DOTs appear to be seeking is an ability to capture, and perhaps measure against benchmarks, the value of their program. It appears that this can best be achieved explicitly or implicitly with a combination of performance measures. Value relies on a combination of many meaningful variables [i.e., value = f (quality, cost, time, etc.)], most or all of which should be tracked to the extent practical.

One example of comprehensive coverage of research program activities by performance measures is from New York. Appendix C provides an excerpt from their Technical Services Division document, "Review of Performance Measures" (*4*). It illustrates a structured undertaking regarding performance measures. A similar example from the Utah DOT is also provided. Virginia's approach to the use of performance indicators is included as well. As will be noted, many of the performance indicators selected by Virginia are quantitative in nature and could lend themselves to the label of performance measure (indeed, the heading of the document indicates the term "measures").

PERCEIVED NEEDS AND PRIORITIES

The questionnaire inquired as to what is currently in use and what needs exist that are currently not satisfied. Responses regarding perceived needs and priorities varied widely, presumably because of the disparity in degrees of implementation of performance measures among the research programs.

For each function, respondents were asked the reason that the function was not currently being measured, and then respondents were asked to identify the primary reason. The following choices were offered:

- Performance measurements are not useful,
- No acceptable method of measurement exists,
- Data are too difficult or costly to obtain,
- Perceived payoff from measures are too low,
- No "mandate from above" is driving the need to measure, and
- Other.

Among states not currently measuring certain performance aspects, the two major program concerns that would benefit from implementation of performance measures were program benefits and post-project implementation. This suggests that if a state DOT is only now attempting to implement performance measures, strong consideration should be given to focusing on these two areas.

The reasons for not measuring performance in these categories were many, but the primary reasons given were reasonably uniformly spread among no acceptable methods of measurement exists, data too difficult or costly, and no mandate from above. These reasons were borne out in follow-up discussions with several of the respondents. For example, in discussions at the 2000 AASHTO Research Advisory Committee national meeting in St. Louis, Missouri, many research managers indicated their desire for a toolbox of proven performance measures that could be readily applied to their programs—the absence of such a toolbox was considered a reason for not currently measuring performance.

Some states, such as New York, indicated some satisfaction with their performance measures but a desire to improve on others.

Several of the respondents reported that their staffing levels or resource allocations were not sufficient to allow for an appropriate job of data collection and analysis. There was even sympathy for this among some of the customers of state DOTs—more than one customer questionnaire was returned with an unsolicited comment about the research unit already being spread as thin as can be reasonably sustained, and that committing resources to performance measures might detract from frontline activities.

A few states (e.g., Louisiana and Virginia) responded that they are not using formal performance measures, but instead using indicators. The implication being that at this time, as a management tool, indicators that can identify trends may be just as effective, and perhaps more cost-effective, than detailed measures and benchmarks. This suggests not only a concern with the cost of data collection, but also of the appropriateness and availability of existing performance measurement methodologies.

Regarding a lack of a mandate from above, Delaware indicated that there appeared to be upper management satisfaction with the program, especially the technology transfer program, and that little or no pressure was being exerted to require performance measurements for the research program. The LTAP center in Vermont also indicated that it is moving toward performance measures (as they are discussed and addressed at a national level), but that requirements beyond those mandated by the FHWA (regarding the minimum number of training programs to be held annually, the provision of a newsletter, etc.) were not being imposed.

Maine had an interesting comment regarding performance measures for post-project implementation—a viewpoint not initially offered by other respondents but concurred with by several during follow-up discussions. Maine noted that there are inherent difficulties in imposing performance measures on the research unit related to the actual implementation of a research project, because the research unit cannot control or require such implementation, as it is under the purview of the operating unit or customer. It is reasonable to infer that there is the occasional research project that is *implementable*, but not necessarily *implemented*, and that there is a question of the fairness of holding the research unit up to a benchmark of implementation without some joint responsibility from the customer unit. On the other hand, there is general agreement among research program staff that their project results should generally be implementable.

CURRENT PRACTICES—PERFORMANCE MEASURES IN USE

In the order of the extent of current usage, performance measures have been implemented among RD&T programs for the following:

1. Project management
2. Post-project implementation
3. Project selection
4. Program benefits
5. Staff productivity
6. Other.

It is interesting to note that post-project implementation performance measures rank second and yet they are identified

as an area needing substantial improvement. Also worth noting is that staff productivity performance measures are usually agency-wide measures. For example, several respondents indicated that the research staff undergoes the same type of annual performance evaluation as do staff in other units. No respondents touted an especially different or meaningful performance measure developed for staff productivity. Because almost all agencies have some sort of staff performance measures through an annual evaluation, the questionnaire responses may be misleading with regard to extent of usage. It seems safe to assume that measures of staff productivity, although probably qualitative instead of quantitative, are in widespread use among state DOTs.

What was clear from the results of the questionnaire is that there are few established benchmarks or goals set for performance measures. This may be because of their relative newness or it may indicate a lack of satisfaction with the performance measures themselves. Where benchmarks are set (either formally or informally), it is also clear that many states take no remedial actions if measures fall below these benchmarks. Although this has no effect on their utility as information providers to the research manager, it does raise questions regarding their usefulness as a management tool.

In the following sections, the actual performance measures reported to be in use by states are discussed, presented in the order in which they were covered in the questionnaire. Not all adhere to the strict definition of performance measures (some are closer to performance indicators), but they are described to provide useful information.

It should be noted that many respondents indicated that their performance measurements are described in their program's transportation research manual. Accordingly, these manuals may be a source of additional useful information for the interested reader.

Performance Measures for Project Selection

There were three types of performance measures used, with two of them being very similar. California mentioned cost-benefit analyses (presumably estimates) as a measure for tracking effectiveness in the project selection process. California establishes benefits compatible with issues of high importance to the DOT. The California Department of Transportation (Caltrans) has a Research Program Advisors Council that reviews the annual slate of projects, which serves to ensure that the selection process meets a qualitative benchmark. Caltrans indicated that they are looking for an improved method to address project selection evaluation. They also initiate a remedial action if project selection does not meet a qualitative benchmark; namely, project leaders are advised of ways to improve the proposed projects in the future. This is an activity that is common to most research units in state DOTs.

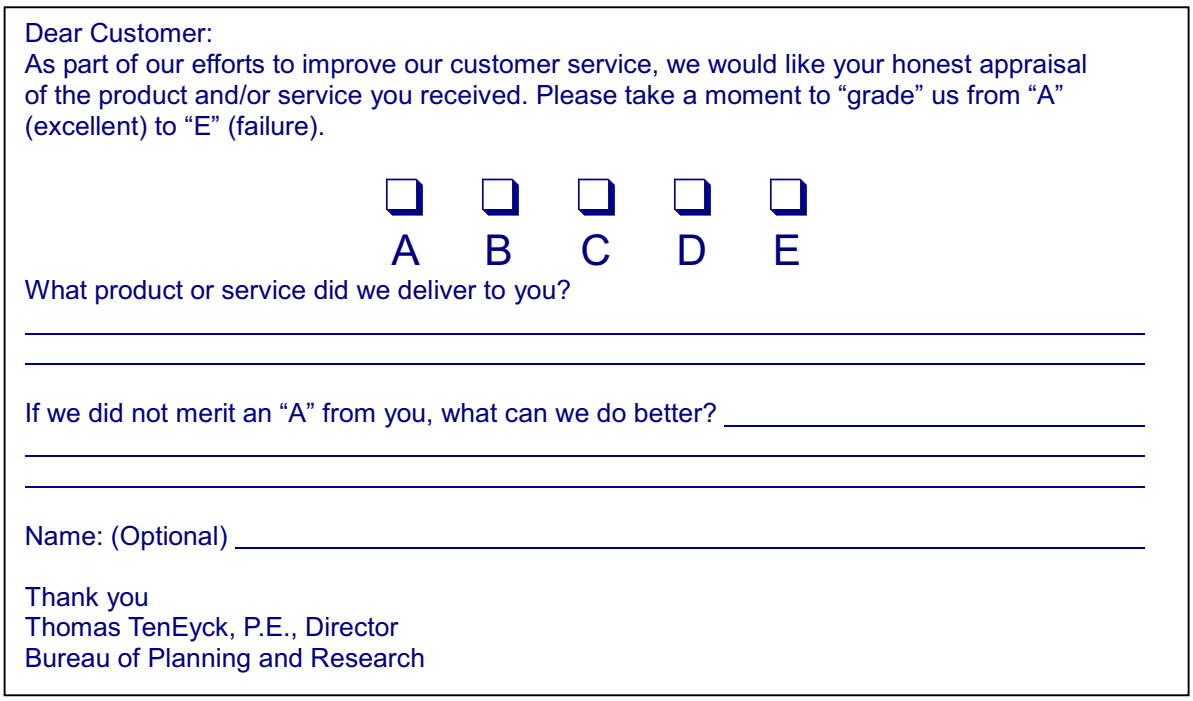
The other two methods described were qualitative alignment of proposed projects with the agency's strategic plan (Colorado) and peer assessment methods (Connecticut, Minnesota, Pennsylvania, and Utah). Implicit in the latter group is that the peer assessment of proposed projects considers the strategic directions of the agency. Pennsylvania also makes use of a customer service index (CSI) card to provide a measurement of how well it is providing the products that are needed. However, the CSI card is generalized—it is not focused specifically on project selection. Nonetheless, such a feedback system should provide qualitative, anecdotal feedback regarding project selection over multiple years, especially if the program is operating at extremes (providing either "all the perfect projects" or "programming none of the needed projects"). The CSI card is shown in Figure 2.

It is not clear from the information gathered as to how in-depth the "needs assessment" efforts of state DOTs were compared to their efforts on actual project prioritization.

Performance Measures for Project Management

Approximately one-third of responding agencies indicated having implemented one or more performance measures related to project management, with many being quantitative in nature. Again, this seems logical, because project management has long required tracking and benchmarking to be effective; long before the term performance measure came into vogue. Essentially, most programs are measuring timeliness (adherence to schedule) and funding (adherence to budget), with the project-by-project benchmark often being the original research contract or project proposal. These are typically being assessed using database methods, and the frequency of measurement is divided into two time periods (with a few exceptions): quarterly and annually.

When asked why these items were measured, varied and useful responses were obtained. Several states (Colorado, Maryland, Minnesota, and Michigan) use timeliness and budget adherence as indirect measures of program and project quality. New York noted that tracking these measures helps to provide information on whether the researchers are doing a good job of planning the research projects. This is an interesting observation in that it provides a measure that could be used to judge the accuracy of future proposals from a specific researcher, and it could be used to provide feedback to the researcher if, for example, the researcher's work were chronically behind schedule. New York commented that it sees delays in the research as not stemming from poor planning or because of legitimate,


Dear Customer:
As part of our efforts to improve our customer service, we would like your honest appraisal of the product and/or service you received. Please take a moment to "grade" us from "A" (excellent) to "E" (failure).

❑ A ❑ B ❑ C ❑ D ❑ E

What product or service did we deliver to you?

_____

_____

If we did not merit an "A" from you, what can we do better? _____

_____

_____

Name: (Optional) _____

Thank you
Thomas TenEyck, P.E., Director
Bureau of Planning and Research

FIGURE 2 PennDOT's customer service index card.

unforeseen delays during experimentation, instead, delays result from failing to consider the effects that new, and as yet, unprogrammed research priorities have on project schedules.

Wyoming noted (and many other agencies probably concur) that they perform qualitative measurements of these matters because the accountability, credibility, and viability of the research program requires at least reasonable adherence to original schedules and funding levels. These are some of the issues described in *NCHRP Synthesis 280* (*5*). It appears to be working, as Wyoming indicated that it is satisfied with this performance measure.

The schedule and funding benchmarks are typically set in one of two ways. The classic way is the schedule and budget established in the research proposal. These become contractual items entered into the database, and progress reports from the contractor provide data for comparison to original estimates. Some states arrive at the final schedule and budget through a negotiation phase between the research unit (sometimes in conjunction with a technical advisory committee) and the researcher.

There is general satisfaction among respondents regarding the performance measures in use, whereas at the same time there was a general sense that researchers' adherence to budget is not much of a problem, but adherence to schedule is. When the benchmark for a measure is exceeded (such as when a researcher is late), most agencies take some corrective action. This corrective action is usually no more serious than notifying the Principal Investigator of the sponsor's awareness of being late. It is clear from discussions among most research sponsors (state, national, and private) that this is a firmly entrenched culture. Most researchers appear to realize that it is difficult if not impossible for a sponsor to provide additional funding (except for unforeseen conditions or for an alteration in the scope of work), but that most sponsors are able, although reticent, to provide no-cost time extensions on research contracts.

Performance Measures for Post-Project Implementation

Approximately 25 percent of the respondents indicated that they have a performance measure in place for post-project implementation. Although some were described as quantitative, all rely somewhat on qualitative information. Minnesota uses a binary process, in which a "closeout memo" is generated after research results are implemented, and the memo is signed by the supporting office director (for the unit that is a client for the research). Accordingly, by tracking the status of closeout memos, Minnesota DOT (MnDOT) can address how well it is getting research results into practice. At the same time as implementation is noted, attempts are made to define the benefits from the research.

In the mid-1990s, the Delaware Transportation Institute (DTI), which performs much of the research for the Delaware DOT, was initiating a similar system. The premise was similar to that of the MnDOT—at the conclusion of a

project, a form would be generated that would require the attention of the director of the operating unit that was a client for a specific research project. The director, based on input from staff, the researcher, and the research unit, would be required to make a formal decision on the status of the implementation of the research results. That decision could include any of the following:

- Research results implemented (and how),
- Research results considered implementable but not implemented (and why),
- Additional R&D required to make results implementable,
- Research results not appropriate for implementation but information of other use (this would cover those useful research projects that discovered that the DOT should not change a specific practice or not use a specific new technology or product), and
- Research results not useful or implementable.

Unfortunately, a formal system was not successfully implemented within the DTI structure, primarily because of a lack of a significant champion for the process and because of a scarcity of available time for consideration of some research results. However, such a system, in a manner similar to that of the MnDOT, helps to make the distinction between the research unit providing results that are implementable (somewhat under the control of the RD&T program) and those that are actually implemented (frequently out of the hands of the program).

The Texas DOT (TxDOT) is in the unusual position of having the state legislature dictate its performance measurement system in this category. The TxDOT evaluates the percent and number of researcher recommendations that are implemented within 2 years of project completion. Goals/benchmarks for this measure are prescribed by the state legislature, but there are no specific actions triggered if performance does not meet the stated goals. The measures are conducted to determine if research results are implementable and how successful the TxDOT is at implementing new products and processes (although there is no way to distinguish whether the research results met the first criterion but failed at the state implementation level). Texas is generally pleased with the approach, but noted that they are trying to change one measure to "percentage of products resulting from research that are implemented within two years." Texas also has a form that denotes how implementation will occur for research results called the "Implementation Plan and Recommendation."

Performance Measures for Program-Level Staff Productivity

The questionnaire results may be somewhat misleading regarding the existence of performance measures for staff productivity, because many states have agency-wide employee performance evaluation procedures in place, and these often address the issue of productivity.

However, several states did indicate the difficulty of adequately measuring the productivity of research staff, given that there are no production units (such as, say, number of design projects completed per year) that lend themselves well to such measures. California noted its difficulty in this area, reporting that no good method had yet been developed. In many cases, the measures were extremely qualitative, with the benchmarks even more so. Utah probably spoke on behalf of many agencies when it responded that its benchmarks for performance measurement in this category were an "educated guess."

Several states did indicate that there were specific actions triggered if performance measures did not meet benchmarks. This characteristic sets this category apart from several others, in which no actions were triggered if performance was below or exceeded benchmark levels. The specific action most commonly triggered by substandard performance was that the employee(s) involved receives a subpar performance evaluation. A smaller subset of respondents also indicated that an employee would receive a correspondingly high rating if benchmarks were exceeded.

Colorado indicated that it is generally not satisfied with the current performance measures, and the belief (perhaps hope) that actual performance relative to benchmarks might someday be rewarded or penalized under a merit pay system.

Performance Measures for Program-Level Benefits

Program-level benefits are often the most important issue a research manager can point to when confronted with the problem of justifying the existence of or investment in the RD&T unit. Accordingly, the need and desire for good methodologies and appropriate performance measures for program-level benefits are compelling. Unfortunately, with the exception of a handful of examples, very few programs have satisfactory performance measures in place.

The leading performance measure in use is cost-benefit ratio, although its application is not consistent from agency to agency. Cost-benefit ratio has the attractive characteristic of being understandable to people outside of the research unit (especially CEOs), of being a similar measure used by other operating units (some planning departments use cost-benefit approaches to make go/no-go recommendations during project development), and at least part of the measure (cost) is reasonably obtained with little special effort.

The fundamental problem with the approach is the establishment of dollar amounts for benefits (especially those associated with savings) or establishing benefits when they are not easily converted to a dollar basis (e.g., environmental improvements). The reasons for the difficulty are many and some are noted here.

- Often, researchers are assigned the task of establishing the benefits, yet they are removed from the implementation and day-to-day understanding of the real costs and savings of the operating unit.
- Any perceived overestimate or indefensible estimate of the benefits of the research program can instantly ruin the entire program's credibility.
- There is no widely used, commonly accepted methodology for assessing project or program benefits, making it difficult to use any values from previous similar projects in a state or values from another state.
- The benefits may represent a finite or infinite series of cost savings or other financial benefit, all of which must be mathematically equated to a present cost so that the ratio may be calculated.
- A basis for assigning dollar equivalents is not commonly accepted for areas such as enhanced safety or environmental improvement.
- Estimates provide quasi-quantitative data, yet are subjective.
- It is hard to capture the true dollar value of politically important projects.

There appears to be agreement among RD&T units that cost-benefit ratio is a generally desirable performance measure, while at the same time there is recognition of the inherent difficulties in its application. These difficulties are similar to those found in the realms of pavement and bridge management systems, where many multiyear optimization programs provide recommendations based on cost-benefit ratios. However, the cost data (and especially user costs) and benefits estimates are gross estimates.

Nonetheless, this performance measure is being used effectively by several agencies, and it is worth consideration by those agencies struggling to decide what to look at for program-level evaluation.

Kansas has a simple and straightforward application of this measure. It is a measure that the upper management at Kansas DOT (KSDOT) determined was "the most effective measure of overall performance of the K-TRAN Research Program." The benefit of each implemented project is determined, and that value is divided by the total cost of the program. Kansas has not established any benchmarks for their benefit-cost ratio, and KSDOT provided a status report from March 2000, indicating that with 33 projects with research results implemented and 26 projects with implementation in progress, the following ratios were calculated:

- Overall benefit-cost ratio: 15.5:1 (to establish this ratio, the numerator includes the benefits of only projects with implemented results and those being implemented, and the denominator is the total cost of the program).
- Benefit-cost ratio on projects with products implemented or being implemented: 26.1:1 (here the denominator is only the cost of projects with implemented results or those being implemented).

The first ratio includes projects that did not result in an implemented product with a determined benefit.

New York takes a somewhat conservative approach to benefit-cost ratio determination and benchmarking. Its entire process is described in its manual of policies and procedures (*16*). First, New York establishes the overall benefits from the program in dollars (including capital/operational savings, user-cost savings, and safety benefits) for completed projects, and it compares them to total research costs. Then New York does two interesting things related to the ratios. First, it establishes a goal of a benefit-cost ratio of 1.0. This may seem very low, but New York notes that by meeting this benchmark based solely from projects, the other services provided by the research program (technical consultation, library/information services) present a clear picture that the research program is providing value to the DOT. Second, New York reports benefit-cost ratios based on a 3-year average. This helps to smooth out spikes and valleys, which is of particular importance when one considers that in 1 year, the results of a particularly successful project may be implemented, resulting in millions of dollars in benefits (possibly on a single large-scale construction project). However, it is not reasonable to expect to have extraordinary projects year in and year out, nor is it desirable to take remedial actions in a program based on a single year's performance. The 3-year averaging is long enough to smooth out spikes and valleys, yet short enough to draw attention should any downward trends occur that require action by management. New York's benefit-cost procedure and a sample calculation are presented in Appendix D. Additional examples are provided in the manual of policies and procedures, to which the reader is referred for step-by-step procedures (*16*).

Florida's research program follows requirements for program-level benefit performance measures that are established at a state agency level by the Office of Policy Planning and Governmental Accountability Transportation Commission.

The shortcomings of benefit-cost measures, especially those related to the "fuzziness" of benefit estimating, can be mitigated through extensive training of personnel and strict adherence to a procedure such as that used by New York. This can allow for program evaluation over time;

however, it will not necessarily lend itself to comparison to other units within the DOT, nor will it lend itself to comparison to other research programs, because not all programs will be working under the same guidelines.

Instead of the quantitative approach of cost-benefit ratios, some states use qualitative measures that are, in effect, peer assessments or user/client surveys. For example, Utah obtains customer feedback through a "letter grading" process. They can then compute an average grade point average on a 4.0 scale that is meaningful to most people. This average is perhaps especially meaningful to any university researchers performing projects for the DOT. This approach, although numerical, is only quasi-quantitative—there are some subjective interpretations required. However, the limited definitions provided as grade descriptions do force the rating scheme to be somewhat consistent among personnel. By providing at least a basic description of what constitutes an "A" project, a "B" project, etc., persons with little understanding of what "successful" research should be can understand the process. Although it is not perfect, it does help different persons to categorize projects similarly, in much the same way that the condition state language that bridge inspectors use helps to provide reasonable uniformity among inspectors looking at deteriorated bridges. A more complete description of Utah's approach is shown in Appendix E, where a customer feedback report from September 1998 is provided. A review of this item provides a sense of the step-by-step process in place for obtaining the customer's views on research. Utah also uses a benefit-cost ratio approach for measuring performance. A report on the ratios for projects and the program is provided in Appendix F.

Several of the step-by-step procedures in use by state DOTs are described in their respective research manuals. These manuals represent a resource for other state DOTs that may be looking to develop step-by-step procedures of their own.

It is worth noting that Florida has recently initiated research that will investigate the costs and benefits of its research projects. The project title is "Review, Analyze and Develop Benefit Cost/Return on Investment Equations, Guidelines, and Variables" (Florida DOT's project reference number is BC353 RPWO #24). The work is being undertaken at the University of South Florida's Center for Urban Transportation Research.

Performance Measures for Other Aspects of Research Programs

Several states responded that they use performance measures to evaluate other functions of the RD&T program. New York, for example, has measures related to technology transfer activities (many other state programs do as well) and to its technical consultation activities. New York noted in their response, as did Delaware in a telephone interview, that getting a handle on more than just the amount of T^2 activity is desirable. The manager of the Delaware DOT's T^2 Center submitted a research project proposal several years ago that proposed looking specifically at how to evaluate the effectiveness of technology transfer activities (the project was not selected for funding).

As noted previously, Pennsylvania attempts to get a general measurement of customer satisfaction through a user/client survey (their CSI). Also, Texas has legislated performance measures related to the number of research institutions under contract and the number of projects under contract. However, as with their legislated performance measures for post-project implementation, there are not specified goals or remedial actions if benchmarks are not met.

In a personal interview, the Vermont Agency of Transportation indicated a few performance measures on which it is working. For example, when a problem arises in an operating unit that requires technical consultation, the Vermont RD&T unit wants to be the first group called. Vermont considers meeting this benchmark as a measure of unit quality. Vermont also noted an interest in setting up performance measures for assessing levels of institutional knowledge versus individual knowledge, as indicators of the potential robustness of programs should retirements occur.

UPPER MANAGEMENT PERCEPTIONS

A separate questionnaire was provided to obtain upper management perceptions regarding the performance of RD&T programs. There was a significant response (22 questionnaires returned, or roughly two-thirds of the number of RD&T program responses), and the responses provided some insights that may be useful to RD&T program managers. These responses are tabulated in Appendix B, after the RD&T program responses.

The first question asked of upper management (frequently a CEO or chief engineer) was whether he or she is familiar with the performance measures used by the RD&T program. Approximately one-third indicated that they were not. Moreover, for those who were familiar with the performance measures, a significant number stated that they did not use the same measures to judge the program. Despite this disconnect in program evaluation, most CEOs provided a positive qualitative rating (6.8 on a scale of 1 to 10) on the performance of their research program. Most of the CEOs had performance measures instituted for operating units in the agency and thus were presumably familiar with them.

One much targeted question asked was what is the one most critical piece of information that the CEO needs to judge the performance of the research program? Although there were a large number of respondents who focused on either the cost-benefit ratio of the program or documented cost savings, there were many who looked for other data (often quantitative) and many of who centered on assurances that the research results were useful (e.g., percentage of projects implemented).

Although CEOs are often pressed for time, it would appear useful for research managers to know the answer to this question. It appears from the upper management responses that this is not consistently done.

Illinois indicated that they are changing the performance measures within the entire agency to be consistent with the "balanced scorecard" approach, a corporate strategic action methodology that has gained recent attention.

CUSTOMER PERCEPTIONS

In addition to obtaining upper management perceptions, a questionnaire was also sent to the customers of the research program within the transportation agency. Although not all states had customers that responded, some states had multiple responses. Again, the questionnaire for the customers is presented in Appendix A, and the responses are tabulated in Appendix B.

As with upper management, customers were asked whether they knew how the research program's performance was measured. The responses suggest that the customers of RD&T do not know and do not care to know how the program is measured as long is it provides value to the agency. The customers also provided a generally high qualitative rating of the research program using a scale of 1 to 10 (overall slightly higher than that reported by the CEOs), providing an anecdotal measure of their aggregated performance.

Customers were also asked to specify their biggest need from the RD&T program. Several provided specific and detailed responses, and such information is something of which most research managers are cognizant but probably should constantly revisit.

One item derived from the customer responses was the usefulness of the library function of the RD&T program. In an age where information is required ever more quickly, it is clear that the operating units within DOTs appreciate the ability of the research program to be a clearinghouse of useful information. This issue may be underrecognized by many research managers, who themselves are adept at finding information in a timely manner. Interestingly, this is an area for which much effort has been placed in the academic arena when it comes to performance measures, and there are various standards or procedures in place to enhance library functions (*12*).

Although not part of the questionnaire, an explicit piece of feedback from customers related to technology training is whether the technology transfer unit gets "return customers." Although not always true, it is a good indicator of a useful program if units in the DOT consistently send their employees to training sponsors by the technology transfer programs in the state.

GAPS (PERCEIVED NEEDS VERSUS PERCEIVED CURRENT ABILITIES/GOALS)

Based on a review of state practice, there are some evident gaps between the status quo and the desired state of operation of RD&T programs.

Before providing specifics, it should be kept in mind that many state DOTs are only now initiating efforts related to performance measures. Accordingly, documents providing substantial background on performance measures could be a most effective learning tool. However, the discussion here will focus on specific gaps that require attention.

Although not specifically reported in questionnaire responses, there is a desire to be able to aggregate the benefits of research from local to state to national levels. The most apparent need related to this is the ability to have apples-to-apples comparisons. The use of dollars would seem to be best suited for such a comparison. However, it is clear that there are no standardized, commonly accepted methods for establishing either the costs or, more dramatically, the benefits from research projects. Although specific states (e.g., New York) do have defensible methodologies, if the methods or assumptions vary from state to state, a credible aggregation of benefits (or an agency-to-agency comparison of cost-benefit ratios) cannot be developed.

Regardless of the desire to be able to provide performance measures that can be accumulated up to the national level, it is clear that most states are not satisfied with their cost-benefit approaches. Primarily, concerns center around the establishment of benefits. The issues with benefits are many, as noted previously. However, this issue must be overcome for there to be useful, credible information over the long term.

Second, alternative performance measures are needed for program-level benefits. It is clear that not all of the activities of the research unit result in value that can be trans-

lated to dollars. The value in educating the future transportation work force through participation in research projects or in technology transfer workshops, for example, is not easily calculated numerically. Nonetheless, there is consensus that some high value does exist. Qualitative performance measures that address these benefits are needed.

Strongly related to the issue of benefits calculation is the need for standardized, commonly accepted performance measures for post-project implementation. Although the benchmarks may vary from agency to agency, it may be very helpful to have established measures used for all research projects that can provide an objective estimate of both the implementability of research projects (the responsibility of the RD&T program) and their actual implementation (a joint responsibility of the RD&T program and the customer).

Along these lines, methods for assessing benefits that have traditionally escaped successful conversion to dollar equivalents are needed. Although some dollar equivalents may be available (e.g., user costs associated with pavement smoothness or traffic congestion that are part of many pavement management systems), they are not universally accepted. This is especially true for issues such as reducing traffic fatalities, where assigning the "value" of human life may be contentious, even if one uses values provided by analyzing payouts resulting from lawsuits instead of interjecting any subjective opinions. Another result from research that would benefit from a "dollar-conversion" approach or a method for capturing value in non-dollar form is enhancement or improvement of the pollution condition resulting from transportation activities.

It is worth noting that whereas many planning activities include cost estimates for construction projects to address improved pavement performance, reduced congestion, etc., it is not clear that the very same dollar values are used in calculating benefits from research that affect those very same areas.

CHAPTER FIVE

OTHER PUBLIC AND QUASI-PUBLIC AGENCIES

Because of GPRA's direct influence on federal agencies, there has been a tremendous amount of work at the national level reviewing what GPRA means to federal agencies individually and collectively. Two reports of the National Research Council looked at the issues of evaluating federal research programs (*15*) and implementing GPRA (*8*). Several of the conclusions and recommendations are noted here, but the reader is referred to the complete reports for additional information. Specifics regarding GPRA requirements are provided in the document, OMB Circular A-11, Part 2, *Preparation and Submission of Strategic Plans, Annual Performance Plans, and Annual Program Performance Reports.*

Regarding the basic issue of evaluation, the first report (*15*) provides the following conclusions:

- Both applied and basic research programs supported by the federal government can be evaluated meaningfully on a regular basis.
- Agencies should match evaluation techniques with the character of their research, taking into account timescales, measurable characteristics, and other items.
- Expert review (quality review, relevance review, and benchmarking) is the most effective means of evaluating these programs.
- A continuing supply of well-educated and well-trained scientists and engineers is needed by the nation.
- Current (as of 1999) approaches for coordinating research programs in multiple agencies with overlapping concerns are insufficient.
- The development of effective methods for evaluating and reporting performance requires input from experts in the field.

The report also provides the following recommendations:

- Research programs should be described in strategic and performance plans.
- For applied research, agencies should measure progress toward practical outcomes.
- For basic research, items such as leadership, relevance, and quality should be assessed.
- Measurement approaches should not be misused (e.g., basic research should not be evaluated relative to short-term relevancy).
- Guidance should be developed to assist in undertaking expert reviews.
- Strategic plans should address the issue of maintaining human resource capacities in fields related to their mission.
- Processes to coordinate overlapping research among multiple agencies should be developed.
- The science and engineering community should play a major role in the implementation of GPRA.

Although targeted at federal agencies, much of this information is pertinent to state and other research programs as well.

A later report (*8*), published in 2001, focused on the implementation of GPRA among the five federal agencies that provide most of the federal funding for research [National Science Foundation (NSF), National Institutes of Health, Department of Defense (DOD), Department of Energy (DOE), and National Aeronautics and Space Administration (NASA)]. Some of the significant findings are noted as follows:

- The agencies had made a good faith effort to comply with GPRA;
- Agencies are using GPRA to improve their operations;
- Reviews by panels of experts have proven to be the most effective technique for evaluating performance;
- Agencies receive differing messages about the desired format, content, and procedures to be used related to GPRA; and
- The use of GPRA results to assist programmatic decision making is not clear.

This conclusion can raise concerns. There is frequently a skepticism that develops within an agency or program if it believes it is being asked to perform routines or provide information, yet that information is perceived to be disregarded during the decision-making process.

Two recommendations from the report are worth noting:

- Federally supported basic and applied research programs should be evaluated regularly through expert review.
- More work on the validation and verification of performance measurement methods is needed.

Again, these recommendations are readily transferable to the state DOT community.

U.S. DEPARTMENT OF TRANSPORTATION

The federal agency with the most profound impact on state DOTs is the U.S. Department of Transportation (U.S.DOT), and at the next level of specialization, the FHWA. The U.S.DOT and FHWA have undertaken efforts related to performance measures for years. It is especially worth noting that the "coordination" activities recommended by the National Research Council reports (*8,15*) have been fostered by the Research and Technology Coordination Council (RTCC), which is administered by TRB. In addition, the DOT Performance Measures Working Group of the RTCC developed a guidelines document, *Setting Goals and Measuring Performance for Transportation Research and Technology Programs* (*17*) in January 1998, to assist organizations in the U.S.DOT with meeting GPRA requirements. An illustrative example from this document is presented in Appendix G.

In May 1999, the U.S.DOT released its *Research and Development Plan* (*18*). This document, in addition to addressing overall strategies for R&D, includes the chapter "Measuring Success," which provides background on GPRA and the U.S.DOT's approach to compliance. Most importantly, it includes some possible performance measures (see Table 1), as well as some possible measures to be used on specific research efforts, including

- percent improvement in accuracy (e.g., in measurement, prediction, and analysis),
- percent improvement in maintainability,
- percent improvement in reliability (e.g., of systems, components, predictions),
- percent increase in capacity,
- percent increase in energy efficiency,
- percent increase in speed,
- percent increase in strength (e.g., for new structural materials),
- percent reduction in emissions or waste products, and
- percent reduction in life-cycle costs.

The U.S.DOT's strategies have been refined, and the goals have been described more explicitly, in a second edition of the *Research and Development Plan* (*19*) published in 2000. Chapter VIII of the updated document, "Measuring the Success of Transportation R&D Investments," provides a useful and detailed account of U.S.DOT activities related to performance measurements.

Work is also underway at the John A. Volpe National Transportation Systems Center on performance measures for transportation R&D. In a presentation at the 2000 AASHTO National Research Advisory Committee (RAC) meeting and a subsequent related paper (*20*), James L. Poage addresses a framework for performance measures for transportation R&D programs. He divides the issue into two parts: technical measures (impacts, benefits, costs of R&D) and programmatic measures (schedule and cost adherence, etc.). He notes that "performance measures should have characteristics that make them useful for decision-making, actions, and communication with stakeholders." The characteristics that he addresses include those

- Driven by decision maker and stakeholder needs,
- Easily understood by decision makers and stakeholders,

TABLE 1
SUGGESTED IMPACT-BASED PERFORMANCE MEASURES FOR TRANSPORTATION (*18*)

Strategic Goal	Impact-Based Performance Measure
Safety	• Reduce speeding-related fatalities 5% by 2000. • Reduce alcohol-related traffic fatalities from 17,000 (1996) to 11,000 (2005). • Reduce motorcycle-related fatalities and injuries 5% by 2000. • Reduce highway fatalities and injuries 20% by 2008. • Reduce child occupant fatalities 15% by 2000 and 20% by 2005.
Mobility	• Provide access to transportation services for all segments of the population within 10 years.
Economic Growth and Trade	• Double the dollar value of transportation-related exports (vehicles, systems, and technologies) within 10 years and triple within 20 years. • Reduce the cost to transport goods and freight by at least 25% within 10 years and 50% within 20 years. • Reduce the time to transport goods and freight by at least 25% within 10 years and 50% within 20 years.
Human and Natural Environment	• Reduce transportation vehicle emissions by a factor of 3 within 10 years and a factor of 5 within 20 years. • Reduce the noise of future transportation vehicles compared to today's inventory by a factor of 2 within 10 years and a factor of 4 within 20 years.
National Security	• Decrease transportation service disruptions due to natural disasters, terrorism, system failure, or other causes by a factor of 5 within 10 years and a factor of 10 within 20 years.

- Reflecting goals and critical success factors of the organization,
- Limited to a critical few, and
- Cost-effective and easy to collect and calculate.

At the FHWA level, much work is also ongoing related to performance measure implementation. The FHWA RD&T unit has developed a workable framework for performance measurement that uses the Baldridge Criteria, which provides a systems perspective for understanding performance management, and aligns with overall agency goals. Work is underway to develop measures of business results to compliment the process-oriented measures that are in place.

In addition, a working paper dated June 28, 2000, "Program Evaluation in FHWA, a New Technique," developed by Clara Conner of FHWA, outlines some thoughts on performance evaluation. This paper focuses on assessing the impact/payoff from an investment in resources. Some of the themes expressed in the paper include

- Objectivity of observation;
- Completeness, reliability, and integrity of data and information; and
- Use of evaluation teams and peer agreement on findings and recommendations.

The paper goes on to note the value in having a systematic, FHWA-wide approach to identifying candidate programs for evaluation and ensuring that programs are evaluated in an objective, organization-neutral manner. This work would be continuously reflected in the FHWA Performance Plan for fiscal year budgets.

The FHWA is also focusing on performance measures related to the partnerships between the FHWA and the state DOTs (personal communication, Connie Yew, April 27, 2001). A report ("FHWA–State Partnership Task Force: Partnership Performance Measures") is being finalized that is related to the Baldridge Criteria for excellence. The report includes the following three key elements related to partnering:

- Program delivery by the FHWA,
- Technical innovation and deployment (including a review of 17 sample technologies that had FHWA and state partnership efforts in deployment), and
- National strategic directions.

The report attempts to determine if the FHWA and state DOTs have similar perceptions on partnering efforts related to technology. For technology deployment and partnering, the FHWA has found that timely response (typically within 1 to 2 days) to state DOTs is a key criterion. Accordingly, timely response may be an appropriate performance measure worth tracking.

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

NCHRP, which acts as the research arm of AASHTO and is administered by TRB, has long been viewed as a successful, high-quality program. NCHRP has not instituted formal, quantitative performance measurements for program-level benefits, although many of the project-level management performance measurements noted in chapter 4 are also conducted by NCHRP. However, NCHRP does have a formal process for collecting and reporting anecdotal and other evidence of program effectiveness. In its periodic status report to sponsors, NCHRP maintains a table of reported uses of its research results, especially uses (such as adoption of recommended specifications) by sponsors and other public agencies. Some of these uses are discovered through serendipity by NCHRP staff, but the reporting of such anecdotal evidence of quality and effectiveness is considered essential to the long-term success of NCHRP. This philosophy is well aligned with the thoughts expressed by Thomas B. Deen, former Executive Director of TRB, at the 2000 National RAC meeting: "Good anecdotes are better than the best benefit-cost data in the world." (This opinion, although held by some, runs contrary to some of the opinions or information derived from questionnaire results and discussions related to this synthesis.)

NCHRP also prepares a similar document for use on special occasions, such as the annual meeting of the AASHTO Standing Committee on Research. It is during this meeting that NCHRP is formally reviewed by its sponsors. In the early 1990s, NCHRP Director Robert J. Reilly instructed NCHRP staff to update a comprehensive table of research outcomes that had specifically addressed AASHTO committee or member department concerns or needs. This table provides an up-to-date, one-stop-shopping answer to the question "What have you done for me lately?" Most of the entries indicate the research project that produced results, the type of result that was provided (e.g., recommended design specifications), the formal action that occurred (e.g., adoption of the specifications by an AASHTO subcommittee), and the party that benefited (e.g., the AASHTO Highway Subcommittee on Bridges and Structures).

At the 2000 National RAC meeting, the Director of the New York RD&T program referred to these very useful approaches as "defensive performance measures." By using these approaches, the unit is able to defend its funding, credibility, accountability, and other resources when confronted about unit activities by upper management, by operating units within the state DOT, or elsewhere.

Although there is no specified benchmark for the number of new entries that should be provided annually, there is a general understanding among staff, and an expectation

among sponsors of NCHRP, that the program will continue to provide useful, readily implementable research results that are of significant value to AASHTO and its member departments.

NCHRP also assesses its performance regarding its processes using survey instruments. For example, the panels that guide NCHRP projects are surveyed three times, covering the project development phase, the active research period (where panel feedback serves to help assess actual versus planned progress), and at a post-project time approximately 4 years after the panel has been disbanded. Respondents are specifically asked to comment on the usefulness of implementation of the research project.

NATIONAL SCIENCE FOUNDATION

In a document available from its website, "Assessing Fundamental Science" (http://www.nsf.gov/sbe/srs/ostp/access/nstcafsk.htm), the NSF presents an interesting approach to the measurement of performance. NSF funds both basic and applied research, and its efforts are credited with some of the long-term scientific and engineering breakthroughs that have had meaningful impacts in the transportation community and elsewhere. These breakthroughs are sometimes so significant that they alter our standard way of performing research.

The document discusses the paradigm shift that occurred based on breakthroughs in computer simulations. Historically, fluid dynamics, and specifically aeronautics, relied on experimentation. However, it was simply not feasible to perform certain experiments that were considered essential to our understanding of, for example, the heating effects of manned vehicles entering the earth's atmosphere. However, over time, these "experiments" could be performed reliably and credibly using computer simulations. As the document describes, the result "was that the space shuttle flew without ever being tested in reentry."

The report provides additional demonstrations of how computer simulation has changed the paradigm of ongoing or planned experimentation (e.g., galactic science, full-scale nuclear testing). The paper also concedes that the paradigm shift has not been complete, noting that drug design is still based "primarily on the experience of the scientist and on [traditional countertop] experimentation."

The NSF document makes the point that such paradigm shifts are an indication of performance. It notes that it is difficult, if not impossible, to predetermine in which fields of study a paradigm shift may occur based on the results of research. This is especially true for basic research. It is also difficult to place a timescale on the periodicity of such paradigm shifts, again because of the nature of basic research. Thus, they are difficult to capture under the moniker of performance measures. Accordingly, the NSF has adopted the use of paradigm shifts as an appropriate performance *indicator*. Because of the impossibility of forecasting such shifts, NSF adopts the strategy of looking backward in time, on a regular basis, to determine what shifts have occurred and when. This will help to determine if the agency's performance is minimally effective or successful. (These two terms are specifically discussed in GPRA—if a program can meet clear criteria for being categorized as one with exceptional timescales for payoff, it does not require performance indicators that are quantifiable and measurable.) However, the NSF has not yet developed any methodological guidance regarding this approach.

Such paradigm-shift investigations are similarly available to state DOTs. The development of computer-aided design and drafting (CADD) has clearly resulted in a paradigm shift regarding the way production work is carried out in the 21st century compared with even the early 1990s. Other such paradigm shifts include the use of reliability theory and limit-states design for structures as a slowly occurring substitute for traditional allowable stress design practices. Certainly, the group or groups within an agency structure that endorsed or promulgated CADD should get some credit for the eventual paradigm shift that occurred.

The NSF also plays another role in the advocacy and promotion of science and engineering in the United States—that of "enabler" for research institutions to have adequate facilities for research and experimentation. To address performance in this area, the NSF chose to look at the following three measures regarding research facilities receiving NSF funding:

- Efficiency of operations,
- Effectiveness of operations to the scientific user community, and
- Effectiveness of activities to the external community.

The basic premise was to "think about performance measures in terms of percentage change from a baseline. The baseline number could be different for each facility, and even measured in different metrics." Some standardization was required, however. The first attempts at this were partially successful, with additional work required on the concept.

In the academic community, the NSF is well known for its "project selection" process. The NSF has always embraced the concept of peer review. Expert panels review proposals, either unsolicited or submitted in response to programmatic and strategic goals of the agency. These panels provide guidance to the NSF in the project selection process, recommending the projects that meet various criteria,

such as quality of the proposal, potential payoff from the research, and human resource development (often measured as the education of graduate students).

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

In the late 1990s, the National Science and Technology Council, which had participation from U.S.DOT, DOD, DOE, the Department of Commerce, the Environmental Projection Agency, and NASA, issued four reports related to transportation technology and research. The reports cut across agency boundaries.

- *Transportation Science and Technology Strategy* (*21*).
- *Transportation Technology Plan* (*22*).
- *National Transportation Science and Technology Strategy* (*23*).
- *Transportation Strategic Research Plan* (*1*).

The first report presented the national strategy (upon which performance would be measured in the future), which was hinged on the following four items:

- Strategic planning and assessment,
- Strategic partnership initiatives,
- Enabling research, and
- Transportation education and training.

Under the heading of "enabling research," the strategy focused "on the long-term evolution of the future transportation system" (*21*). Six research areas were identified that required federal support, because it was perceived that any single private company could not reasonably be expected to recover investments (and thus it was anticipated that sufficient private sector RD&T response in these areas would not result).

- Human performance and behavior;
- Advanced materials;
- Computer, information, and communication systems;
- Energy and environment;
- Sensing and measurement; and
- Tools for transportation modeling, design, and construction.

Four areas were noted under education and training.

- Introduction of transportation concepts: elementary and secondary education,
- Vocational and technical training,
- Transportation degree programs: international and multidisciplinary, and
- Mid-career transportation training.

Goals and benefits of the 10 areas above were developed, but performance measures were not included in the strategy report.

The next report (*22*) went one step further by prescribing specific outcomes in various vision areas from the original document. For example, under the umbrella vision area of "accessibility for aging and transportation-disadvantaged populations," the report specifies a desired outcome of "deployment of welfare-to-work transportation strategies in all 50 states by 2003." Under the vision area of "transportation and sustainable communities," a target outcome is to "minimize the adverse impacts of transportation projects on wetlands and, where impacts are unavoidable, replace 1.5 acres of wetlands for every 1 acre affected."

How closely results come to target outcomes could certainly be a performance measure, but the report itself does not address the issue of measuring performance.

The third and fourth documents (*23,1*) are simply updates of the first, but they include some of the specific outcome elements (revised as necessary) from the second document as well. However, they remain relatively silent on the issue of how to measure performance.

These four documents are critical to the issue of performance measurement, however, because of the previously stated importance of aligning at least some of an agency's performance measures with the strategic plan of the agency.

OTHER AGENCIES—INTERNATIONAL

The United States is not alone in its current focus on performance measurements for research programs. In 1996, a report entitled *Performance Evaluation Mechanisms for Transportation Research Programs* was prepared for the Transportation Association of Canada (*24*). This document was a synthesis of practice primarily in Canada, although it also touched on national U.S. practice (e.g., TRB and FHWA) as well. The main finding was that "most . . . organizations are carrying out relatively little formal evaluations of their R&D. Most of the evaluation is conducted at the beginning of the R&D process . . . [there is] little evaluation of the R&D after it has been completed."

Where evaluations were being used to assess performance, the following categories were noted:

- Expert opinion
- User/client opinion
- Cost-benefit methods
- Case studies
- Performance indicators.

Although the document had little information readily implementable regarding performance measures, it does serve as an excellent primer for any U.S. agency that is only now embarking on their efforts.

The report does describe a system for assigning relative priorities to broad research fields. The system is used by the Commonwealth Scientific and Industrial Research Organization in Australia and is a graphically based system. A similar system is used by many U.S. researchers as part of specific research projects when determining where to focus their research efforts.

Essentially, the method involves assigning a numerical but qualitative ranking of the attractiveness and feasibility of various research projects or areas, then graphing them in two-dimensional space. Areas or projects that lay furthest from the origin and close to a line representing a 45 degree angle represent probable good areas for investment. This technique is shown in Figure 3, which is a combination of two figures from the Canadian report.

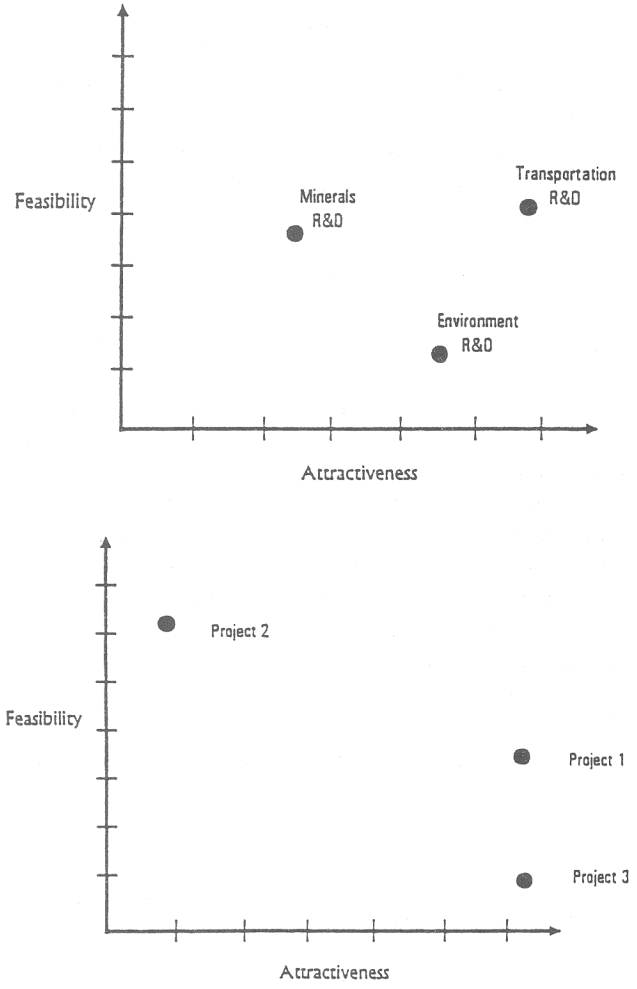


FIGURE 3 CSIRO project ranking method (*24*).

The report contains a summary table showing the author's perceptions of the applicability to various phases in the RD&T process. This table is included as Appendix H. The report also provides some background information on user/client surveys and on cost-benefit evaluation; however, this information is very general and provides no detailed discussion in comparison with that provided in Appendix D.

A recent paper (*25*) focusing on technology policy describes some of the tools and methods available for assessing technology programs. The paper includes discussions of both U.S. and European programs, and its insights into European programs may be useful to U.S. research managers. The report stresses the evaluation of the socioeconomic impacts of research in universities and public laboratories, as well as other issues that are gaining prominence in the United States (e.g., evaluation of collaborative R&D). The report discusses various earlier studies, and includes a description of the three-part classification of evaluation methods.

- "Retrospective, historical tracing of the knowledge inputs that resulted in specific innovations.
- Measuring research outputs in aggregate form from particular sets of activities . . . using bibliometrics, citation accounts, patent counts, compilations of accomplishments, and so forth.
- Economic theory/econometric methods employing, as measures of performance, productivity growth, increase in national income, or improvements in social welfare as measured by changes in consumer and/or producer surpluses."

The report summarizes performance evaluations for several European organizations or agencies, such as the Alvey Programme (intended to revitalize the United Kingdom's information technology sector through RD&T) and the European Union Framework Programmes (a broad-based pan-European initiative related to RD&T). One interesting finding noted was the "outcome" of doctoral training because of European investment in research projects. This reinforces the theme reported in the United States of the importance and success of human resource development as a product of research investment.

The report provides a very high-level discussion of performance evaluation and provides a useful insight for high-level management or perhaps national or federal agencies. The information, although interesting, does not seem to translate directly into tools that can be readily implemented by state DOTs or other similar research programs.

CHAPTER SIX

PRIVATE SECTOR EXPERIENCES

There has been a long-held suspicion that private sector RD&T programs must have the most viable and precise performance measures and measuring methods, because they are "turning a profit" based on their work. Although the information collection efforts for this synthesis were not as successful in the private sector as originally hoped, it is clear that the suspicion is not necessarily valid. As there is among public sector agencies, there is variation in private sector practice, from lack of formal performance measures to intricate ones.

DIFFERENCES BETWEEN PRIVATE AND PUBLIC SECTORS

The most, or perhaps only, significant difference between the private and public sectors is that most performance measures in the private sector are driven on a dollar basis (be it cost-benefit, profitability, or whatever). The limited information reviewed for this synthesis found little emphasis being placed on the importance of "developing the next generation of transportation (or other) professional," whereas this is considered a useful and strategic goal in the public sector. The perceived needs and priorities of the private and public sectors are similar when it comes to accountability, credibility, and program justification.

CURRENT PRACTICES—PERFORMANCE MEASURES IN USE

Current practice in the private sector indicates a range of activities similar to those in the public sector. Several examples are discussed here. It is clear that the "silver bullet" that many in the public sector may be searching for does not necessarily exist among corporate RD&T programs; however, there is still meaningful information to be gained from a review of private sector practices. It is worth noting that whereas most public sector agencies function in a world where sharing information, practices, and strategies is commonplace, the private sector is characterized by firms that exist across the spectrum, from those that share in a manner similar to public agencies, to those that share little or are unresponsive. Given the competitive and fast-paced nature of private sector business, this should not be considered a negative characterization, only an indication of what one has to be prepared for if one attempts to solicit additional information from corporate interests.

It is clear that cost-benefit or dollars-based performance measures are *de rigueur* in corporate America. R&D in the private sector is essentially in the same class as venture capital in general—high risk is assumed, yet high payoff is expected. Three examples of performance measure uses in the private sector are described here. As can be seen, the approaches are not unique to the private sector.

Cost-Benefit Ratios

The General Motors Research Laboratory (GMRL) demonstrates a classic example of the use of cost-benefit performance measures. As described by W. G. Agnew (presentation, 2000 National RAC Conference, St. Louis, Missouri), GMRL use a measure known as a "harvest ratio" and it is based on

- Projects completed during the 3 previous years (similar to New York's time frame);
- Projects with cash benefits over $500,000; and
- Benefits documented by the customer.

In a single snapshot of GMRL activity covering about 1,000 projects

- 76 were promising candidates;
- 16 were documented successes;
  - 8 had shown cost savings,
  - 4 had resulted in product improvements,
  - 3 had resulted in warranty reductions,
  - 1 had resulted in quality improvement.

GMRL defines the harvest ratio (H) as follows:

$$H = \text{(dollar benefits)/(GMRL annual budget)}$$

In the mid-1980s, harvest ratios of around 0.75 were reported.

GMRL then uses the harvest ratio to calculate a return on investment. This performance measure is well understood through the corporate management structure of GM. The calculation of return on investment is shown here.

$$\frac{NPV}{I} = H\,1 - \frac{(1+k)^{-n}}{k} - 1$$

where

NPV = net present value,
I = investment,
H = harvest ratio = 0.75,
k = discount rate, and
n = number of years.

Setting $NPV/I = 0$

k_0 = internal rate of return = 72%
(corporate hurdle rate for investment = 15%).

Fluid-Share Approach

A business practice that became attractive with corporate downsizing can serve as a pseudo-performance measure related to staff productivity. The practice is called the "fluid share" in Australia. It is based on the assumption that given the opportunity work forces will request additional resources (e.g., additional staff) unless there is a direct incentive to each participant to minimize resource use. The fluid-share concept is based on profit sharing. Suppose that a program unit has a staff level of 20 persons who share equally in net profits. Then suppose that a downsizing effort is desired by management—reducing the staff size to 16 persons without changing the production expectations (and thus overall profitability resulting from production). In most cases, members of the 16-person work force would clamor for additional staff. However, if the work force is offered a choice—share all the profits among only 16 members or hire an additional one or two staff and then share the profits among 17 or 18, many work forces will opt for working at the downsized level. However, this management style also recognizes that work forces know their own limitations. If a staff of 16 is insufficient, then quotas will not be met, with the result that overall profits for the group are threatened. In such a case, the work force itself may opt to request additional staff or resources, knowing that they are, in effect, optimizing individual profits.

Translation of this practice to a public sector agency is not simple, but there are potential applications where certain resource allocations might be made available instead of profits. How the system optimizes itself could lead to a performance measurement, perhaps by looking at overall resource requirements of groups or the agency as a whole. Many potential pitfalls exist (including infighting for staff members perceived to be more productive), and thus careful consideration should be given to this idea before it is pursued in any depth. It is, however, an innovative technique that includes strategic goals such as empowerment of the staff that are addressed in *NCHRP Synthesis 280* (*5*).

User/Client Surveys and Peer Assessments

Synthetic Industries (SI) is a manufacturer and distributor of products that are used both inside and outside of the transportation industry. For example, they are one of the largest producers of carpet backing. They are also involved with geosynthetics and concrete reinforcement. SI aggressively develops and funds ideas that it believes have the potential to become solutions that can be used by the marketplace. Its R&D staff includes approximately 20 trained engineers, whose efforts are tied to the corporate goals of expanded production, increased sales, tighter efficiencies, and higher quality products.

Regarding performance measures for quality, SI relies on internal and academic researchers, as well as obtaining customer feedback through surveys and focus group sessions. These performance measures are qualitative, with the benchmark being set as "exceeding their expectations." SI has generally been satisfied with these performance measures (personal communication, Karen Baker, May 31, 2001).

Interestingly, SI also touts its role in human resource development, although it does not measure performance in that regard. It actively seeks out academic partners for research projects and assists in technology transfer efforts. As with any private sector activity, however, there are undoubtedly some marketing goals that are included in this function.

Trinity, a manufacturer of roadside safety and other apparatuses, has a straightforward use of de facto customer assessments to measure the performance of its development efforts: If a product gets into use, it is considered "quality" (personal communication, D. Johnson, June 5, 2001).

GAPS

From the limited information available, it appears that the fundamental gap in knowledge for performance measures for the private sector mimics one from the public sector: the quantification of benefits from RD&T efforts for use in cost-benefit analyses. However, one can surmise that their benefit data may be somewhat better on average because of the "hard numbers" with which the customer units deal every day.

CHAPTER SEVEN

ACADEMIC SECTOR EXPERIENCES

The academic sector provides much of the work force involved in research projects and technology transfer efforts sponsored by state DOTs and other public and private agencies. Accordingly, it is worthwhile to investigate how this sector measures its performance, and to compare and contrast those measures with those of the sponsors. This chapter attempts to describe the differences between the academic and public sectors, define some of the current measures in use in academia, and provide insights as to where the two systems are not well aligned.

DIFFERENCES BETWEEN ACADEMIC AND PUBLIC SECTORS

A primary difference between the academic and public sectors is the nature of what is delivered by each. The public at large accepts that academic institutions teach and generate knowledge. They deal in the merchandising of enlightenment and intellectualism. The public at large is relatively unaffected in the short term by successes and failures in academia. At best, a good academic experience prepares a student for an enriching career and life. At worst, the student has spent some time and money to learn more about him or herself. The public tends to accept that measuring the performance of academic institutions is a very difficult thing to do—they get daily samplings when they read in the newspaper or see on the television the difficulties of assessing the outcomes of K–12 education. However, K–12 education is something that most people are familiar with, whereas many have little concept of what constitutes a college education. Without a firm grasp of what is supposed to be delivered, there is little room for criticism. It therefore comes as no surprise that the primary public criticism of academia is based on dollars—namely, skyrocketing tuition costs, a measure that the public understands and appreciates. Accordingly, it should not be surprising that the sensitivity to performance measures within the academic sector differs substantially and meaningfully from those used in the public or private sectors, especially when it comes to the various products or processes being measured.

The public sector, on the other hand, either provides tangible goods or needed services. The public takes notice when a road is not built on time, or when some public service is not being provided as promised. Although an exceptionally small percentage of the public knows how to construct a road, many have an intuitive feeling as to what needs to be done to build one. This is in direct contrast to academia, where the abilities of a professor are sometimes viewed with "awe" (deserved or not), simply because that individual holds a doctorate or works in a highly technical field of study. These differences help to set the stage for the different performance measures that exist in the academic versus the public sector.

PERCEIVED NEEDS AND PRIORITIES

To understand the RD&T activities of academia, one must first consider the broader realm in which academics operate and the philosophies that many espouse. This includes views on education in general, as well as their awareness of outcomes-based evaluation, such as that done for programs and curricula through the Accreditation Board for Engineering and Technology. Furthermore, one must recognize that many, and perhaps most, academics perform research of some level on a routine basis, although its significance in the overall duties of a professor will vary drastically depending on items such as type of institution, position in the promotion and development track, etc. In addition, academics are involved with great frequency in technology transfer, be it in traditional college courses or in special workshops or seminars. Thus, almost every aspect of their professional lives is related to the subject of this synthesis.

Academics, in general, are concerned with quality and productivity. However, the concept of "value" is not frequently coupled with quality. Other than the establishment of the list of eight schools known as the "Public Ivies" (public schools that provide an educational and research experience supposedly on par with the Ivy League institutions), which specifically touts the combination of quality and lower cost, there is no frequent discussion among college faculty or administration of the cost-benefits of their activities.

Academics *do* tout the benefits of a college education to students, such as in career earnings increases. However, rarely is such discussion brought to the level of specific dollar-related measures of performance for individual professors.

Thus, the academic world is continually searching for performance measures that provide a true indication of quality, but that do not necessarily address the economic constraints under which the public or private sector performs.

CURRENT PRACTICES—PERFORMANCE MEASURES IN USE

For this synthesis, several academic institutions were contacted to acquire some basic information. Although the queries were targeted to specific institutions, the following questions were generally asked; all specifically targeted toward research activities by faculty:

- What performance measures, if any, does the institution use to *quantitatively* track quality and value?
- What performance measures, if any, does the institution use to *qualitatively* track quality and value?
- Does the institution have a good idea of the performance measures that are used by *sponsors*? (In other words, do the faculty have a good grasp of how NSF, DOD, and other research sponsors measure the performance of their research programs?)
- Are there any performance measures typically used in academia that seem to be in place for historical or traditional reasons, but are not necessarily *good* indicators of actual performance?
- If an institution's measures indicate that benchmarks are not being met, is there any plan for corrective action to help meet the benchmarks in the future?

Many of the measures used are quantitative and historical, but questions remain regarding their effectiveness at assessing performance, especially as it relates to *applied* research. Some of these methods are discussed here. The first few deal with individual performance, whereas the last is concerned with institutional/organizational performance. Although many faculty performance evaluation techniques have been in existence for a long time, it is an area ripe for research, especially as academia becomes increasingly exposed to calls for accountability. In 1990, an overview of faculty performance appraisal techniques and research was prepared outlining some new thoughts on faculty evaluation (*13*).

Publication Quantity and the Science Citation Index

The generation of scholarly articles has long been known as one of the key performance indicators for young faculty or faculty desiring promotion. Even the general public is familiar with the phrase "publish or perish!" The performance of faculty members is often largely based on the number of papers accepted for publication by scholarly journals. The expectation is that, because of the peer review process typically employed in the acceptance process, quality and the impact of the work is assumed if publication is to occur.

Many in the academic community have used a specific performance measure to address the impact of one's research output. The Science Citation Index is a database of journal articles. The philosophy behind the index is that a relatively small number of journals publish the bulk of *significant* scientific results. Accordingly, journals are ranked (they are given what is sometimes referred to as "impact factors") based on their significance, which is based on peer review, timeliness, and other items. Thus, to have a paper published in a journal with a high impact factor would potentially be perceived as a higher level of performance than publishing in a journal with a low impact factor. (One problem with this is that many practitioner-oriented transportation-related journals, including those of the TRB, have low impact factors, thus diminishing the incentive for high-quality faculty to publish in them. This may simply be an issue of academia being slow to embrace different types of publications that can still be scholarly, but not in the traditionally accepted sense.)

Student Evaluations

Many college graduates remember the brief time they spent each semester evaluating their professors' performance. These evaluations can be considered part of the institutional performance measures for personnel assessment related to technology transfer, although they are primarily qualitative (although a numerical score is often provided instead of such words as good, fair, or poor). The importance of these evaluations in the promotion and tenure process varies, with more importance being placed on them at so-called teaching institutions in comparison to research universities. Indeed, the question is often asked if student evaluations should be discontinued (*13*). However, it is commonly accepted among faculty that while the measure is flawed at best, it does provide some useful anecdotal information that can be used to improve a course or delivery style. Similar evaluations are used when faculty teach technology transfer courses (such as through workshops) to practitioners.

Dollars

Many research institutions have a performance measure of "dollars generated per faculty member." Few institutions have established benchmarks (being above the average is, of course, an often used, self-imposed goal of a professor) or specific remedial actions that take place if a benchmark is not met. One problem with this approach is that it is very insensitive to the costs of performing different types of research. Large-scale experimental research on bridge construction many cost many times that of a transportation policy study. Thus, in the "dollars generated" column, the

former seems quite impressive. However, the impact from the less expensive study may far outweigh the bridge study, and this is not explicitly accounted for in the measure. Clearly, the consideration of the generation of dollars is not an outcomes-based approach.

Peer Assessment

One of the performance measures embraced by academics is peer assessment. This is primarily confined to research proposals, but also has its place in employment evaluation, such as in the promotion and tenure process. Peer assessment appears to be the most accepted method for evaluating quality and importance, and the number of peer-assessed and funded projects does provide some valid insight as to the quality of the researcher.

It should also be noted that at many institutions, faculty members are given performance credit for simply submitting a proposal to a sponsor, regardless of its final disposition. This signifies a significant departure from public or private sector practice, where some results are required. In essence, the current practice is like taking home part of a win just for showing up at a baseball game! However, one must realize that the writing of proposals is, in effect, a professional development activity for a professor. Extending the baseball analogy, one would not expect the typical player to hit a home run the first time he swings the bat—the same holds true regarding expectations for funding the first time a professor writes a proposal. Practice makes perfect.

Organizational Performance Measures

Although still somewhat insulated from the economic and political realities faced by the public and private sectors, there have been increasing calls for accountability in the academic sector.

In Texas, some performance measures are dictated by the state legislature. The Texas Transportation Institute (TTI) prepares an annual report for the legislature that addresses specified outcome measures, projected performance related to that measure, and actual performance. If projections and actuals are significantly different, an explanation of the variance is provided.

Some examples of outcome measures or outputs from the *Annual Report on Measures: Fiscal Year 2000* (*26*) include

- Dollar volume of research studies performed,
- Leverage ratio of general revenue appropriations to total funds,
- Number of TTI patented safety devices installed,
- Number of students involved in TTI education and research activities, and
- External dollars brought to the state of Texas.

Vermont Technical College underwent self-examination with an objective of generating performance measures for strategic activities and goals at the department and institutional level—these goals are aligned with classroom instruction, research, development, and technology transfer roles of faculty. The intention was to generate a continuous improvement tool based on performance. It was called the Assessment-Analysis-Adjustment (A-A-A) program. The effort started with each department determining what performance measures to assess (in one of the engineering technology curricula, for example, this included, for those students who chose to take the exam, the percentage of students who passed the Fundamentals of Engineering examination, which is the first step toward licensure as a professional engineer). Many additional performance measures were determined by departmental faculty.

The intent was to create a set of benchmarks for each measure. For the example shown previously, the benchmark was to meet or exceed the national average for engineering technology programs. As data were collected each year, the next step would be to analyze the data and compare them to benchmarks, being cognizant of "noise" in the system or other external factors that could affect these data. Finally, decisions are made to adjust the departmental operations to help meet or exceed benchmarks in future years.

On paper and in theory, the A-A-A program held great promise. It suffered, however, from the same fate as many performance-measurement attempts; trying to collect too much data without sufficient determination of which data are the most significant. Accordingly, the A-A-A program is being revised and implemented, with hopes that it will eventually be more manageable and therefore embraced by the faculty and college as a whole.

University Transportation Center Performance Measures

University Transportation Centers (UTCs) are funded by the U.S.DOT and have research and technology transfer functions. Reporting requirements for the UTCs are specifically spelled out (*27*) and are a contractual matter for continued funding. Interestingly, the performance indicators (the term used in ref. *27*) are quantitative and very similar to the historical numbers-based evaluations done at universities. For example, the following are considered performance indicators for the UTCs, although some of

them have tenuous ties to the RD&T activities traditionally seen in government transportation departments:

- Number of courses offered;
- Number of academic departments offering them;
- Number of students completing the courses;
- Number of students involved in transportation research projects;
- Number of students enrolled;
- Number of students receiving degrees;
- Number of transportation research project proposals submitted to center;
- Number of transportation research projects awarded by center;
- Total budgeted costs for those projects;
- Number of individuals listed as principal investigators in those projects awarded;
- Number of visitors to transportation center website;
- Number of transportation seminars, etc., conducted for practicing professionals;
- Number of UTC newsletters published;
- Number of peer-reviewed transportation research reports and books published;
- Number of transportation papers accepted for presentation at academic/professional meetings; and
- Number of external awards received for transportation research.

This list is not exhaustive. It does indicate, with the exception of the last few entries, that there is little focus on quality or value of the performance of the UTC, but instead an assurance that certain levels of activity are maintained. This is not in and of itself a bad thing. However, it is clearly not sufficient for the long-term improvement of UTCs or for the comparison of the performance of different UTCs.

To their credit, some UTCs consider the performance measures shown above to be a baseline and are moving forward with the development of more quality- and value-based measures.

Libraries

Except for their location (on an academic campus instead of in government offices), the issue of the performance of academic libraries is much the same as for government/DOT libraries. As mentioned previously, an outstanding resource for anyone responsible for the performance of libraries is *Measuring Academic Library Performance: A Practical Approach* (*12*). This document was prepared for the Association of College and Research Libraries and contains a general discussion of library systems, some user/client satisfaction surveys, facilities usage rate information, and guidance in analyzing performance data.

GAPS

With the previous discussion as background, it is now acceptable to provide some specific examples of practice with respect to the five questions posed to academic institutions for this synthesis. The responses help to identify some gaps requiring additional attention. A review of these responses makes it clear that there is a disconnect between the world that the academic researchers and teachers are working in and that of the state DOT or other RD&T programs.

- What performance measures, if any, does the institution use to *quantitatively* track the quality and value of the research performed by faculty?

In 2000, the Center for Transportation Research and Education (CTRE) in Iowa indicated that they undertook a campaign to identify program benefits. Some anecdotal information was found, but nothing that allowed for quantitative measures. In essence, the best that was achieved was the ability to "detect success," but not necessarily to be able to tabulate it

In an attempt to semi-quantitatively track performance, CTRE Director Stephen Andrle noted that an indicator of success is the evolution of research into services. Specifically, he relayed the story of geographic information systems (GIS) technology. Iowa DOT funding of research and development allowed the agency to become involved in GIS before they were prepared to staff it internally. Now the Iowa DOT has a competent and growing applications group in this area. Andrle further noted that if a service becomes something that someone will pay for, this is a true measure of performance.

David Fowler, Director of the International Center for Aggregates Research (ICAR, a joint venture of the University of Texas at Austin and Texas A&M University), noted that ICAR has no formal list of measures. He indicated the great difficulties in trying to quantify quality, especially for research whose payoff may not be known for many years. ICAR does track some quantitative performance measures, including number of publications, oral presentations to professional groups, and leveraging of funds.

TTI has an interesting list of performance measures, in addition to those prescribed by the state legislature. TTI's Director, Herb Richardson, considers repeat business from sponsors (a likely surrogate measure for quality of the work performed), research dollars per faculty, economic benefits from implemented projects, and peer reviews. Although TTI does not have a very structured way of tabulating these results, they do prepare a series of 1-page documents called "Returns on Research." These serve as a kind of performance measure.

In the College of Engineering at the University of Delaware, some very traditional performance measures are used. According to Dean Eric Kaler, the quantitative measures in place include dollars raised, students educated, papers published, and the quality or impact of those papers.

- What performance measures, if any, does the institution use to *qualitatively* track the quality and value of the research performed by faculty*?*

The qualititative measures vary greatly, with "rewards for excellence" being among them. ICAR noted that customer satisfaction is a key qualitative measure, with a statement of "You did a good job" from a sponsor being a meaningful indicator of performance. Other qualitative measures are similar to those used in DOT RD&T programs.

- Does the institution have a good idea of the performance measures that are used by the *sponsors* of research? (In other words, do the faculty have a good grasp of how NSF, DOD, and other research sponsors measure the performance of their research programs?).

The overwhelming response to this question is that faculty members do not know or understand the performance measures that are used by their sponsors. It is also clear that a one-time statement regarding what the measures are is not sufficient. CTRE noted that there are two cultures clashing: that of client-driven research (e.g., state DOTs) and that of researcher-driven research. This lack of understanding or appreciation on the part of academics is not necessarily a bad thing; that is, it does not fundamentally undermine the quality of the research. However, it does probably have a significant influence on issues such as timeliness of the research or the need for implementability of the research. Thus, it can affect the *value* of the research.

This lack of understanding by faculty members reinforces the importance of framing research questions in a way that will most likely be understood. It also suggests that the question, "Does the complete research address the customer's need?" can be part of the formulation of a performance measure.

There were some concerns that the project management (process management) performance measure reporting requirements sometimes imposed by research sponsors were so overwhelming and stifling that they adversely affected the quality of the research and the opportunity for breakthrough findings. There seems to be an acceptance of some levels of accountability, but strong concerns that the Principal Investigator's attention and funding are directed toward reporting business measures instead of toward the actual research.

- Are there any performance measures typically used in academia that seem to be in place for historical or traditional reasons, but are not necessarily *good* indicators of actual performance?

There is agreement that the number of papers published alone is not a good metric. Some of the measures used by transportation center directors at universities are very client-oriented. These differ from the historic measures (number of graduate students, number of papers published, dollars brought in) that tend to be more strongly embraced by academic deans at various colleges within an institution. This disparity in appreciation for types of measures mimics the disparity in the career paths of these two groups. Center directors are increasingly non-doctorates who have a long career in practice, sometimes with a stint as a research sponsor. Academic deans tend to follow a more traditional path of professor to department chair to dean. Accordingly, their measures of research performance align with their life experiences.

- If your institution's measures indicate that research is not meeting benchmarks, is there any plan for corrective action to help meet the benchmarks in the future?

The typical response is that future funding for specific faculty is jeopardized, but that no institution-wide measures are typically initiated.

As one can see, there are significant differences in perception and practice between those that sponsor the research and those that perform it in an academic setting. There is frequently a philosophy of either "I am paying the bills, so the researcher must come around to my way of thinking" or, alternatively, "I know what is best for the research, so I will focus on that and tend to the sponsor's concerns only if there is time." Performance measures that indicate quality and value and that fall somewhere in the middle are presumably the best requirement.

Although academic institutions are slow to change, this does not mean that they will not change. As the world becomes more fast-pace, so do they, just as their counterparts in state RD&T programs. A measured ability to adapt to change rapidly may be a performance measure of the future for academic institutions. To quote Graham Wallis, a professor at the Thayer School of Engineering at Dartmouth College, who was interviewed in the Spring 2001 issue of the school's "Directions" magazine and has overseen (as dean, senior professor, and in other roles) decades of change at what is generally perceived to be a responsive institution to change and sponsor needs, "Running Thayer School was like driving a sports car as opposed to a bus . . . ."

CHAPTER EIGHT

CONCLUSIONS

Since the publication of the Canadian synthesis, *Performance Evaluation Mechanisms for Transportation Research Programs* in 1996, little has occurred in state DOTs that had not already embraced performance measures at that time. Currently, there is an increased interest in and attention to such measures; however, their use is sporadic and varies among transportation agencies. There is no commonly accepted set of performance measures, either qualitative or quantitative, that can be unequivocally recommended as the baseline set that an organization should institute. Furthermore, although there is a strong desire among many DOTs to pursue performance measures, there is no uniform agreement on the need for them. Many DOTs do not have agency-wide mandates for their implementation. Some states are experimenting with them as a management tool.

Despite a scarcity of specific information or quantitative practice on performance measures for research, development, and technology (RD&T) programs, a few common threads run through the experiences of state DOTs, national organizations, and the private and academic sectors. These commonalities hint at the best practices that may be employed for assessing performance.

- Ties to strategic goals

Most performance measures that appear to be working or seem satisfactory to the organization using them have ties either directly or indirectly to the strategic goals of the department.

- Measures that are understandable to upper management

Upper management is comfortable with concepts such as cost-benefit ratio or having anecdotal evidence of the quality and utility of research. Although there exists an awareness of its shortcomings, cost-benefit approaches appear to be very common and very useful to those organizations that have implemented them.

- Relative simplicity

Even for those agencies applying benefit-cost ratio approaches there is relative simplicity and elegance in their methods. An extremely refined approach is not used where there is an acceptance of uncertainty in the data (e.g., benefits in dollars). The calculation procedures do not become overwhelming and thus they become easier to implement uniformly among staff.

It should come as no surprise that there were no mentions of linear regressions, fifth-order equations, multivariate statistical analyses, or similarly high-level functional operations among those measures that seem to be performing well.

- Cost-effectiveness of data

Most organizations seem aware that the use of less but more significant data is important unless there is appreciable improvement to the understanding of performance from the efforts to collect, analyze, and make decisions based on additional data. Moreover, many organizations make effective and extensive use of data that they generate for the standard business operations.

- Utility and usefulness of peer assessment, especially for qualitative measures

It is clear that among all sectors (more so public and academic than private), peer assessment is considered a useful method of performance measure. Although it may not provide quantitative outputs, it can provide quasi-quantitative results, because of the averaging effects of groups of people confronted with normative qualitative scales. Many organizations that use peer assessment to provide performance feedback are perceived to be among the most vital and robust programs—these include the NSF, NCHRP, and many of the state DOT RD&T programs that are flourishing.

- Usefulness of the human-resource development component of RD&T

Most of the state, federal, private, and academic sector organizations contacted for this synthesis indicated that human resource development is a useful and necessary outcome of RD&T activities. Accordingly, performance measures that account for this are among the best practices that can be instituted.

If one looks at the types of performance measures that are being used, and the comments of those using them or in search of useful measures, the following additional findings become evident:

- Of all of the functions in RD&T programs, the strongest need for performance measures seems to be for program-level benefits.

- Moreover, there is a need for quantitative measures for program-level benefits that can provide longitudinal information on the performance of a program, as well as within-agency and agency-to-agency comparisons of the research program with other programs.

- A close second regarding the need for performance measures is the area of post-project implementation. This need seems to have two parts: a measure that indicates how much implementation occurs and a measure that indicates the payoff from implementation. Minnesota's "closeout memo" may serve as a model for states determining how to best address the issue of the extent of implementation. Measures of the payoff from implementation vary greatly and are not rigorous or robust.

- Concerns exist regarding performance measures for project implementation, given that the actual implementation of research results can be outside the control of the RD&T program. There seems to be a desire for two measures related to implementation: one that indicates whether the research is implementable and another that indicates whether it was actually implemented.

- Peer or expert dialogues or assessments have been documented as being among the most effective tools for evaluating performance. This is stated in national studies, and the concept has been embraced and lauded by state DOTs in both the ongoing RD&T Peer Exchanges (where the focus is on information exchange and dialogue, not assessment) and the up-and-coming local technical assistance program (LTAP) Peer Exchanges. Peer assessment has been the mainstay of agencies such as the NSF and NCHRP.

- Performance measures that are tied to the strategic goals of the transportation agency tend to obtain upper management support.

- Upper management and customers of research programs are not concerned with how the RD&T program measures its performance, but instead focus on whether the program meets their expectations. A need exists for enhanced coordination and understanding between the involved parties.

- The library function of the research unit should not be undervalued; its usefulness is noted by customers of the research program.

Because of the variations in practice within the United States; because of the very different degrees to which RD&T programs have studied, developed, or implemented performance measures; and because embarking on such implementation can be costly, some conclusions can be drawn from the information presented herein that may provide a useful context for considering future actions within transportation agencies. It may be worthwhile for RD&T programs to consider these issues as they develop or improve their performance measures, or even make the decision whether or not to initiate their use.

- Managers of RD&T programs may be most successful by initially developing a limited number of performance measures, focusing on those that both most strongly indicate performance (quality, productivity, etc.), while also being cost-effective to implement.

- Consideration should be given to tying performance measures for RD&T programs to the current strategic goals of the overall agency, but not exclusively. By tying such measures to strategic goals, the relevance of the program becomes more evident to upper management. However, given that strategic goals within a department can change with each incoming administration (and even more frequently), the RD&T program may also require performance measures that can provide longitudinal indications of how well the program is working. The RD&T program may want to work to make these long-term measures part of the agency's strategic goals from administration to administration.

- Research is needed to provide guidance on the application of cost-benefit ratios for assessing program-level and project-level benefits of research programs. Such guidance need not be "one size fits all," but it is clear that many variations exist among research agencies, with no clear superior method. Special emphasis could be placed on techniques for estimating benefits from research that can feed into the cost-benefit analysis. The use of multiyear averaging to smooth out spikes and valleys appears to have merit.

- Research should be initiated to address benefits that are known to be important, but are virtually impossible to quantify. For example, research could be undertaken to assess the value of a research project that shows that a process/technology should not be implemented, because such findings are generally considered useful. In addition, specific methods for addressing non-dollar benefits (e.g., safety improvements, environmental enhancements) could be investigated or developed.

- Managers of RD&T programs may want to embrace the collection, use, and analysis of anecdotal information, or perhaps even the development of full-blown case studies, as a qualitative performance measure.

These measures may provide a quantifiable supply of answers to the question: "What have you done for me lately?"

- Closely aligned with the previous recommendations is the need for research to develop performance measures that can be aggregated to various levels and across state and sector (public, private, academic) boundaries. To justify expenditures in RD&T activities in the United States, it is important that credible assessments of the utility of research activities be made. Furthermore, it is important that credible benchmarks be determined that can set the stage for true enhancements in program management to exceed those benchmarks.

- Research managers may want to consider the development of performance measures that address the following fundamental questions:

  – Are the research results implementable?
  – Have the research results been implemented?

Success with performance measures related to these two questions will require input from customers, researchers, and others. These issues should be separated to allow for customer ownership and responsibility for implementing research results that are considered implementable. In many cases, the customer unit may be able to justify additional resource allocations to overcome the initial barriers associated with such implementation.

- Research managers may wish to consider performance measures that relate to the following three primary concerns of the unit and to the activities required to improve unit performance over time:

  – Process management
  – Program quality
  – Program value

- RD&T programs should remain cognizant of developments in the education arena for two reasons: (1) much of their research work force (university researchers) live their daily lives in this arena, and (2) education assessment deals with measuring intractable and intangible outcomes such as "learning." Advances in performance measurement techniques in the realm of education may be readily transferable to RD&T.

- Consideration should be given to the continued development of performance measures, such as the efforts spurred by TRB Committee A5001 (Conduct of Research), for technology transfer that goes beyond the traditional measurement of extent of activity. Stronger indications of the effects of training are needed. Efforts by groups such as TRB Committee A5012 (Technology Transfer) and the national LTAP organization should be given full support by member departments (for those states that have an active role in funding and guiding LTAP activities) and others (in states that do not have a state DOT-based LTAP program).

- Although potentially difficult to institute within a state DOT setting, consideration should be given to the use of NSF's backward-looking investigation for paradigm shifts as a critical and strong indicator of RD&T program performance. Such an approach can provide a clear signal as to how research affects the agency.

- Finally, research managers and other personnel may want to consider taking courses on "corporate physics," if available. Such courses are often taught at management-friendly engineering schools or technology-friendly business schools. Corporate physics refers to the study of business practices in a quantitative and scientific way, with many analogies drawn between the behaviors of social systems (e.g., business organizations) and their counterparts in the physical realm (e.g., demonstrations of control strategies that work for first-order management issues based on the same mathematical analyses used in first-order physical systems). These courses can provide additional information on performance measurement, the mathematics of policy and process decisions, and the application of statistics to quasi-technological systems.

REFERENCES

1. National Science and Technology Council, *Transportation Strategic Research Plan*, Office of Science and Technology Policy, Cambridge, Mass., 1999.
2. Poister, T.H., *NCHRP Synthesis of Highway Practice 238: Performance Measurement in State Departments of Transportation*, Transportation Research Board, National Research Council, Washington, D.C., 1997.
3. "Primer on Performance Measurement," U.S. Office of Management and Budget, Washington, D.C., 1995. [Online]. Available: http://www.c3i.osd.mil/bpr/bprcd/4130.htm.
4. "Review of Performance Measures," *Operational Goal 94-8*, Technical Services Division, New York State Department of Transportation, Albany.
5. Deen, T.B. and B.T. Harder, *NCHRP Synthesis of Highway Practice 280: Seven Keys to Building a Robust Research Program*, Transportation Research Board, National Research Council, Washington, D.C., 1999.
6. Cambridge Systematics, Inc., *NCHRP Report 446: A Guidebook for Performance-Based Planning*, Transportation Research Board, National Research Council, Washington, D.C., 2000.
7. *Webster's New Collegiate Dictionary*, G & C Merriam Co., Springfield, Mass., 1961.
8. *Implementing the Government Performance and Results Act for Research: A Status Report*, Committee on Science, Engineering, and Public Policy, National Academy of Sciences, National Academy of Engineering, Institute of Medicine, Washington, D.C., 2001.
9. Friedman, M., *A Guide to Developing and Using Performance Measures in Results-Based Budgeting*, The Finance Project, Washington, D.C., 1997.
10. Research and Technology Coordinating Committee, *Special Report 256: Managing Technology Transfer: A Strategy for the Federal Highway Administration*, Transportation Research Board, National Research Council, Washington, D.C., 1999.
11. Phillips, G.W., *Technical Issues in Large-Scale Performance Assessment*, National Center for Education Statistics, U.S. Department of Education, Washington, D.C., 1996.
12. Van House, N.A., B.T. Weil, and C.R. McClure, *Measuring Academic Library Performance: A Practical Approach*, American Library Association, Chicago, Ill., 1990.
13. Gabbin, A.L., S.N. Cairns, and R.L. Benke, Jr., *Faculty Performance Appraisal, Center for Research in Accounting Education*, James Madison University, Harrisonburg, Va., 1990.
14. Brown, M.G., *Keeping Score: Using the Right Metrics to Drive World-Class Performance*, Amacon, New York, N.Y., 1996.
15. *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*, Committee on Science, Engineering, and Public Policy, National Academy of Sciences, National Academy of Engineering, Institute of Medicine, Washington, D.C., 1999.
16. *A Manual of Policies and Procedures for Operation of the Transportation Research and Development Bureau*, Transportation Research and Development Bureau, New York State Department of Transportation, Albany, N.Y., 1997.
17. *Setting Goals and Measuring Performance for Transportation Research and Technology Programs*, Research and Technology Coordinating Council, Transportation Research Board, National Research Council, Washington, D.C., 1998.
18. *U.S. Department of Transportation's Research and Development Plan*, U.S. Department of Transportation, Washington, D.C., 1999.
19. *U.S. Department of Transportation's Research and Development Plan*, 2nd Ed., U.S. Department of Transportation, Washington, D.C., 2000.
20. Poage, J.L., "Developing Actionable Performance Measures for Government Programs," *Journal of Cost Management*, March/April 2001.
21. National Science and Technology Council, *Transportation Science and Technology Strategy*, Office of Science and Technology Policy, Washington, D.C., 1997.
22. National Science and Technology Council, *Transportation Technology Plan*, Office of Science and Technology Policy, Washington, D.C., 1998.
23. National Science and Technology Council, *National Transportation Science and Technology Strategy*, Office of Science and Technology Policy, Washington, D.C., 1999.
24. Williams, D., *Synthesis of Practice No. 4: Performance Evaluation Mechanisms for Transportation Research Programs*, Transportation Association of Canada, Ottawa, 1996.
25. Georghiou, L. and D. Roessner, "Evaluating Technology Programs: Tools and Methods," *Research Policy*, Vol. 29, No. 4, 2000.
26. Texas Transportation Institute, *Annual Report on Measures: Fiscal Year 2000*, Texas A&M University, College Station, Tex., 2000.
27. Research and Special Programs Administration, *Reporting Requirements for University Transportation Centers (UTCs)*, U.S. Department of Transportation, Washington, D.C., 1998.

APPENDIX A

Questionnaires

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM
Project 20-5, Topic 31-04

Performance Measures for Research, Development, and Technology Programs

Questionnaire Sent to State DOT RD&T Programs

Many transportation research, development, and technology (RD&T) programs have recognized the value of using performance measures to track their effectiveness. This questionnaire is being conducted to gather information for a synthesis that will document the kinds of performance measures in use, explain how they were developed, describe their effectiveness, and determine the common themes and best practices among programs.

The information that you supply with this questionnaire will be very helpful in establishing the state of the practice as well as future needs. Please return your completed questionnaire, along with any supporting documents, **by June 30, 2000**, to:

Scott A. Sabol
Architectural & Building Engineering Technology Department
Vermont Technical College
PO Box 500, Main Street
Randolph Center, VT 05061

or by FAX to 802-728-1390 (return via mail is preferred)

If you have any questions, please call Scott Sabol at 802-728-1272, or e-mail him at *ssabol@vtc.vsc.edu*

Below, please provide the information requested for the person completing the questionnaire or whoever should be contacted to obtain any follow-up information. Several questions are intentionally brief, and information that is more detailed may be sought through a telephone or e-mail follow-up.

Name:_____
Title:_____
Agency/Organization:_____
Mailing Address: _____
_____
Telephone:_____ FAX:_____ E-mail:_____

Thank you for your assistance in completing this questionnaire! Please go to the following page.

The following explanation of terms may be helpful as you complete the questionnaire:

Performance measure: An objective appraisal, usually quantitative, of something that indicates accomplishment toward a goal. For example, if a goal is for the research program to complete most projects on or under budget, then a performance measure for that goal may be "percentage of projects that are completed on or under budget."
Quantitative: A numeric measure of amount, number, or extent. For example, "75% on-time completion of research projects" is a quantitative measure.
Qualitative: A non-numeric measure indicating traits, characteristics, or trends. For example, "Identification of high-priority projects by the research program is judged to be good to very good" is a qualitative measure.
Research: Investigation or study, usually requiring creative thought instead of the application of existing knowledge or tools. Research is often characterized as basic, applied, or a mixture.

Basic: No definitive final application of the research results is anticipated or guaranteed (the quest for knowledge for the sake of generating the knowledge, with the hopes of its future usefulness).
Applied: An application of the research results is foreseen, and the anticipated or desired result can often be described before the research is undertaken or completed.

Development: Advancement or enhancement of an idea or concept, often through creation of a tangible good.
Technology: Equipment, tools, skills, or knowledge, often used to attempt to meet an objective.
Technology transfer: Training, explaining, or enabling others in the use of equipment, tools, skills, or knowledge.
Effectiveness: How useful or helpful something is with respect to meeting a goal.
Efficiency: How quickly or how little effort is required with respect to meeting a goal (Note: a process can be efficient (it may be simple) but not effective (it may not provide meaningful progress toward meeting the goal)).
Project-level: Information, measurements, characteristics, and similar items that have to do with a specific project. For example, the on-time completion of a single project, or the cost savings associated with the product of a single project, are project-level measurements.
Program-level: Information, measurements, characteristics, and similar items related to an operational program as a whole. For example, the average on-time completion of all projects in a research program is a program-level issue, as would be the perceived contribution by the research program to a state DOT's effort to achieve its mission.
Project programming/development: Identifying or scoping an individual RD&T project.
Program development: Identifying the full slate of RD&T projects, plus all other RD&T activities, that define the role and activities of the RD&T program.

INSTRUCTIONS: Please mark answers as appropriate (e.g., with a checkmark), and/or write in information as appropriate. You may append additional pages or send supplementary material as well. Please write your organization/agency name on the top of each page (in case pages from your response get separated).

1. **Does your organization have a formal mechanism to measure the performance of its RD&T program?**

___ Yes ___ No

(If your answer is NO, you may skip to Question 4, Page 7)

2. **If YES, for which of the following categories do you measure performance?**

___ Project selection/programming (how well is the program picking important projects?)
___ Project-level management (are individual projects on time, on budget, and delivering the expected products?)
___ Post-project implementation (are completed projects put to use?)
___ Program-level staff productivity (are the type and amount of results that are expected being produced?)
___ Program-level benefits (what effect is research unit activity having on the transportation system?)
___ Other categories (please describe):

3. **For each category in Question 2 that you <u>do</u> measure, please answer the following questions:**

(a) ***Project selection/programming*** *(Picking the "right" projects to do)*

(i) Is the performance measure: ___ qualitative? ___ quantitative? ___ both?
(ii) Is there a written description of the performance measure? ___ Yes ___No
(iii) How is performance measured? (describe or attach written policy)

(iv) How frequently is performance in this category measured?

___ annually ___ every project selection cycle___ every "X" years, where X = _____
___ other (describe)

(v) Why is this category measured? (why is it important?) (describe)

(vi) How are the goals/benchmarks for performance in this category established?

(vii) How frequently are the performance goals updated?
___ annually ___ every two years ___ with every strategic or long-range plan
___ other (please describe)

(viii) Are any actions triggered if performance does not meet goals? ___ Yes ___ No ***(IMPORTANT NOTE: If you mark yes here <u>or to similar questions below</u>, Scott Sabol will contact you later for details, or you may send an attachment that provides additional explanation.)***

(ix) Are any actions triggered if performance exceeds goals? ___ Yes ___ No

(x) Is your organization satisfied with the performance measures used for this category? Please describe, and provide any other comments on project selection/programming.

(b) ***Project-level management*** *(Are projects on time, on budget, and delivering the expected products?)*

(i) Is the performance measure: ___ qualitative? ___ quantitative? ___ both?
(ii) Is there a written description of the performance measure? ___ Yes ___No
(iii) How is performance measured? (describe or attach written policy)

(iv) How frequently is performance of this category measured?
___ annually ___ every project selection cycle___ every "X" years, where X = _____
___ other (describe)

(v) Why is this category measured? (describe)

(vi) How are the goals/benchmarks for performance in this category established?

(vii) How frequently are the performance goals updated?
___ annually ___ every two years ___ with every strategic or long-range plan
___ other (please describe)

(viii) Are any actions triggered if performance does not meet goals? ___ Yes ___ No

(ix) Are any actions triggered if performance exceeds goals? ___ Yes ___ No

(x) Is your organization satisfied with the performance measures used for this category? Please describe and provide any additional comments on project-level management.

(c) ***Post-project implementation*** *(Are completed projects put to use?)*

(i) Is the performance measure: ___ qualitative? ___ quantitative? ___ both?
(ii) Is there a written description of the performance measure? ___ Yes ___No
(iii) How is performance measured? (describe or attach written policy)

(iv) How frequently is performance of this category measured?

___ annually ___ every project selection cycle___ every "X" years, where X = _____
___ other (describe)

(v) Why is this category measured? (describe)

(vi) How are the goals/benchmarks for performance in this category established?

(vii) How frequently are the performance goals updated?
___ annually ___ every two years ___ with every strategic or long-range plan
___ other (please describe)

(viii) Are any actions triggered if performance does not meet goals? ___ Yes ___ No

(ix) Are any actions triggered if performance exceeds goals? ___ Yes ___ No

(x) Is your organization satisfied with the performance measures used for this category? Please describe and provide any additional comments on post-project implementation.

(d) ***Program-level staff productivity*** *(Are the type and amount of results that are expected being produced?)*

(i) Is the performance measure: ___ qualitative? ___ quantitative? ___ both?
(ii) Is there a written description of the performance measure? ___ Yes ___No
(iii) How is performance measured? (describe or attach written policy)

(iv) How frequently is performance of this category measured?
___ annually ___ every project selection cycle___ every "X" years, where X = _____
___ other (describe)

(v) Why is this category measured? (describe)

(vi) How are the goals/benchmarks for performance in this category established?

(vii) How frequently are the performance goals updated?
___ annually ___ every two years ___ with every strategic or long-range plan
___ other (please describe)

(viii) Are any actions triggered if performance does not meet goals? ___ Yes ___ No

(ix) Are any actions triggered if performance exceeds goals? ___ Yes ___ No

(x) Is your organization satisfied with the performance measures used for this category? Please describe and provide any other comments on program-level staff productivity.

(e) ***Program-level benefits*** *(What effect is research unit activity having on the transportation system?)*

(i) Is the performance measure: ___ qualitative? ___ quantitative? ___ both?
(ii) Is there a written description of the performance measure? ___ Yes ___No
(iii) How is performance measured? (describe or attach written policy)

(iv) How frequently is performance of this category measured?
___ annually ___ every project selection cycle___ every "X" years, where X = _____
___ other (describe)

(v) Why is this category measured? (describe)

(vi) How are the goals/benchmarks for performance in this category established?

(vii) How frequently are the performance goals updated?
___ annually ___ every two years ___ with every strategic or long-range plan
___ other (please describe)

(viii) Are any actions triggered if performance does not meet goals? ___ Yes ___ No

(ix) Are any actions triggered if performance exceeds goals? ___ Yes ___ No

(x) Does your organization select or use performance measures that will assist it in "telling its story?" (That is, using performance measures to demonstrate the need to have an effective/active research program.)

(xi) Is your organization satisfied with the performance measures used for this category? Please describe and provide any other comments related to program-level benefits.

(f) ***Other category*** (please describe the category if you complete this section—if none, go to Question 4 on next page):

(i) Is the performance measure: ___ qualitative? ___ quantitative? ___ both?
(ii) Is there a written description of the performance measure? ___ Yes ___No
(iii) How is performance measured? (describe or attach written policy)

(iv) How frequently is performance of this category measured?
___ annually ___ every project selection cycle___ every "X" years, where X = _____
___ other (describe)

(v) Why is this category measured? (describe)

(vi) How are the goals/benchmarks for performance in this category established?

(vii) How frequently are the performance goals updated?
___ annually ___ every two years ___ with every strategic or long-range plan
___ other (please describe):

(viii) Are any actions triggered if performance does not meet goals? ___ Yes ___ No

(ix) Are any actions triggered if performance exceeds goals? ___ Yes ___ No

(x) Is your organization satisfied with the performance measures used for this category? Please describe and provide any other comments.

4. **Which categories of your program are currently <u>not</u> being measured with respect to performance but you believe <u>should</u> be? After each that you check, please circle <u>all</u> the applicable reasons from choices A–F below, then the <u>single</u> primary reason).**

___ Project selection/programming (Circle all applicable reasons: A B C D E Primary reason: A B C D E)
___ Project-level management (Circle all applicable reasons: A B C D E Primary reason: A B C D E)
___ Post-project implementation (Circle all applicable reasons: A B C D E Primary reason: A B C D E)
___ Program-level staff productivity (Circle all applicable reasons: A B C D E Primary reason: A B C D E)
___ Program-level benefits (Circle all applicable reasons: A B C D E Primary reason: A B C D E)
___ Others (please describe):

A - Performance measurements are not useful
B - No acceptable method of measurement exists
C - Data are too difficult or costly to obtain
D - Perceived payoff from measure too low
E - No "mandate from above" driving the need to measure
F - Other reasons (please describe here, indicating which category you are discussing):

5. **Does your organization use performance measures for units other than the research unit? (e.g., preconstruction, maintenance and operations, planning)** ___Yes ___ No ___ Don't know

6. **If the answer to Question 1 was NO, how does your organization evaluate its effectiveness? Please describe:**

_____
_____
_____
_____

7. **What is the single most important thing necessary to help you better measure the performance of the research program?**

8. **Does your organization differentiate performance measures among research activities, development activities, and technology (or technology transfer) activities?** ___ Yes ___ No

If yes, how?

9. **Supporting information:**
___ Check here if supporting documentation is being provided with this questionnaire response.
___ Check here if your organization has an organization-wide policy related to performance measures.
___ Check here if you are providing your organization's most recent "FHWA Peer Exchange" report (this is encouraged!).

10. **Any final comments?**

Thank you for completing the questionnaire. Please return your response, preferably by mail instead of FAX, no later than <u>June 30, 2000</u> to the address shown on the first page.

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Project 20-5, Topic 31-04

Performance Measures for Research, Development, and Technology Programs

SPECIAL QUESTIONNAIRE FOR STATE DOT UPPER MANAGEMENT

Many transportation research, development, and technology (RD&T) programs have recognized the value of using performance measures to track their effectiveness. A detailed questionnaire has been sent to your agency's RD&T program to gather information for a synthesis that will document the kinds of performance measures in use, explain how they were developed, describe their effectiveness, and determine common themes and best practices among programs. For the purposes of this questionnaire, the following definition is pertinent:

> ***Performance measure:*** *An objective appraisal, usually quantitative, of something that indicates accomplishment toward a goal. For example, if a goal is for the research program to complete most projects on or under budget, then a performance measure for that goal may be "percentage of projects that are completed on or under budget."*

To help to obtain information about upper-management perceptions on the effectiveness of the RD&T program that relate to performance measures, your input is requested on this brief, two-page questionnaire.

Please return your completed questionnaire **by June 30, 2000**, along with any supporting documents, to either your RD&T program manager (who will compile additional parts of the questionnaire and send a coordinated response) or directly to:

Scott A. Sabol
Architectural & Building Engineering Technology Department
Vermont Technical College
PO Box 500, Main Street
Randolph Center, VT 05061

or by FAX to 802-728-1390 (return via mail is preferred)

If you have any questions, please call Scott Sabol at 802-728-1272, or e-mail him at *ssabol@vtc.vsc.edu*

Below, please provide the information requested for the person completing the questionnaire or whoever should be contacted to obtain any follow-up information. **This special questionnaire should be completed by upper management within the DOT**, preferably at a level that can observe the effects of the RD&T program on the overall transportation system in your state.

Name:_____
Title:_____
Agency/Organization:_____
Mailing Address: _____
_____
Telephone:_____ FAX:_____ E-mail:_____

Thank you for your assistance in completing this questionnaire! Please go to the following page.

1. **Are you familiar with how your RD&T program measures its own performance?** ___ Yes ___ No

   If yes—please *briefly* describe how the program measures its performance:

2. **Do you measure the RD&T program's performance in the same manner as it self-evaluates its performance?** ___Yes ___No

   If No, what items/issues do *you* track to evaluate the RD&T program?

   Do you have goals/benchmarks for the RD&T program to meet? ___ Yes __ No

   Are specific actions triggered when goals/benchmarks are met or not met? ___ Yes __ No

   If yes, what are those actions?

3. **On a scale of 1 (low) to 10 (high), circle the number that best matches your perception of the RD&T program's effectiveness at improving transportation within your state (through cost savings, improved safety, discovery of innovative practices, etc.):**

   1 2 3 4 5 6 7 8 9 10

4. **What is the most critical information that the RD&T program can provide you with as a measure of its effectiveness?**

5. **Does your agency use quantifiable performance measures for units <u>other than</u> the RD&T program?** ___ Yes __ No

   If Yes, which units? (e.g., construction, maintenance, planning)

Thank you for completing the questionnaire. Please return your response, preferably by mail instead of FAX, no later than <u>June 30, 2000</u> to the address shown on the first page.

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Project 20-5, Topic 31-04

Performance Measures for Research, Development, and Technology Programs

SPECIAL QUESTIONNAIRE FOR RESEARCH PROGRAM CUSTOMERS

Many transportation research, development, and technology (RD&T) programs have recognized the value of using performance measures to track their effectiveness. A detailed questionnaire has been sent to your agency's RD&T program to gather information for a synthesis that will document the kinds of performance measures in use, explain how they were developed, describe their effectiveness, and determine common themes and best practices among programs. For the purposes of this questionnaire, the following definition is pertinent:

> ***Performance measure:*** *An objective appraisal, usually quantitative, of something that indicates accomplishment toward a goal. For example, if a goal is for the research program to complete most projects on or under budget, then a performance measure for that goal may be "percentage of projects that are completed on or under budget."*

To help to obtain information about perceptions on the effectiveness of the RD&T program among the customers of the research program, your input is requested on this brief, two-page questionnaire.

Please return your completed questionnaire **by June 30, 2000**, along with any supporting documents, to either your RD&T program manager (who will compile additional parts of the questionnaire and send a coordinated response) or directly to:

Scott A. Sabol
Architectural & Building Engineering Technology Department
Vermont Technical College
PO Box 500, Main Street
Randolph Center, VT 05061

or by FAX to 802-728-1390 (return via mail is preferred)

If you have any questions, please call Scott Sabol at 802-728-1272, or e-mail him at *ssabol@vtc.vsc.edu*

Below, please provide the information requested for the person completing the questionnaire or whoever should be contacted to obtain any follow-up information. **This special questionnaire should be completed by a customer of the state DOT's research program (e.g., an end user of research products).**

Name:_____
Title:_____
Agency/Organization:_____
Mailing Address: _____
_____
Telephone:_____ FAX:_____ E-mail:_____

Thank you for your assistance in completing this questionnaire! Please go to the following page.

1. **Are you familiar with how your RD&T program measures its own performance?** ___ Yes ___ No

   If yes—please *briefly* describe how you believe that the program measures its performance:

2. **Do you measure the RD&T program's performance in the same manner as you think it self-evaluates its performance?** ___Yes ___No

   If No, what items/issues do *you* track to evaluate the effectiveness of the RD&T program?

3. **On a scale of 1 (low) to 10 (high), circle the number that best matches your perception of the RD&T program's effectiveness at improving transportation within your state *by providing useful research products for you and your staff:***

   1 2 3 4 5 6 7 8 9 10

4. **Does the RD&T program contact you to determine the outcome or "payoff" from implemented research results?**

   ___ Routinely—they contact us in a systematic manner
   ___ Occasionally, but not in a systematic manner
   ___ Rarely or never

5. **If the RD&T program contacts you regarding the "payoff" from research, how often are you able to provide feedback (preferably quantitative) regarding the usefulness or outcome from using the research results?**

   ___ Routinely—we attempt to characterize the payoff in a systematic manner
   ___ Occasionally
   ___ Rarely or never

6. **What is the single most important thing that the RD&T program could do (or already does for you) to help you with your unit's activities?**

7. **Does *your* operating unit have performance measures (with established benchmarks/goals) against which it is regularly evaluated?**

Thank you for completing the questionnaire. Please return your response, preferably by mail instead of FAX, no later than <u>June 30, 2000</u> to the address shown on the first page.

APPENDIX B

Summaries of Questionnaire Responses

STATE DOT RD&T PROGRAMS

QUESTIONS 5–8 AND 10

STATE	5	6	7	8	10
	PMs in other units	If not PM, how evaluate unit	Most important need or measuring issue/tool/info	Differentiate among RD&T	Other comments
AK					
AZ	Y	Meeting goals established during strategic planning	Desire by upper management for PMs	Construction-related projects measured by on-time performance	
CA			Useful methods with wide acceptance		
CO	Y		Tracking of deployed research	Y–through individual performance plans	
CT			Completion of quarterly program reviews	Y–some tailoring	
DE	Y		Need full-time research unit with staff	Y–some special PMs used for T^2 activity	
FL	Y		Awareness of need	N	
GA	N		Staffing to allow the performance to be adequately studied	N	
HI	N				
IL			Feedback from customers	N	New PM system expected by end of 2001
KS			Calculation of benefits from implemented research		Reports published, distributed, etc., also tracked
KY	Y	Timeliness & implementation	Effective implementation of research results	N	
LA	Y		Develop meaningful indicators	N	
ME	Y	DOT strategic plan, goals, and objectives. Also “transportation indicators”	Institutionalizing the implementation of research results across MDOT		See below
MD	Y		Usefulness of research	N	
MN					
MS	Y		Input from construction & maintenance units to determine if findings are put into practice (and to identify problems for future research)	N	
MT			More time/staff	N	
NE	Y	Periodic informal checks & reviews of completed research	Tracking and documenting implementation of research results; identifying the success factors of a research program	N	
NV	N	Related to meeting goals set in Peer Exchange process	Agreed-upon, acceptable methods of measurement with buy-in from top management and technical divisions (“customers”)	N	
NH					
NJ	Y	See below	Customer satisfaction and research implementation surveys	N (may later)	

NY	Y		Accurate quantification of benefits (done in conjunction with client to avoid appearing self-serving)	Y (no benefits assigned to tech transfer activities)	See below
OK	Y		Feedback on how the information provided through research was used and if it was useful in the decision-making process	N	See below
PA	Y		Customer input	Y See below	
RI	N	See below.	Information from other units on how research results are used	N	
SC			Establishment of fair and realistic methods of measuring performance		
TN		See below	Additional staff and input from customers	N	
TX	Y		Standard methodology for estimating program-level benefits	N	
UT	Y		Good data	Y (different goals, end users, and feedback mechanisms)	
VA					
WY	Y		Buy-in at the executive-staff level to all the staffing to increase to allow for adequate performance measurement	N	

ME: Some performance measures were developed years ago (e.g., no. of projects completed, no. of recommendations given to a policy group), but these have not been implemented nor are they expected to.
NJ: As a new organization, the first years are being spent organizing—-measuring of performance comes later.
NY: Without performance measures, the true worth of the research program to the agency is never recognized. This is okay in good times, but in bad times, the research program must have the numbers to defend its continued presence.
OK: Efforts that require too much labor are not effective. Customer satisfaction is a driving concern, so many efforts are geared toward that end.
PA: T^2 services are measured after each class and a biennial survey is conducted.
RI: RIDOT is a small organization and its effectiveness can be done through observation (without formal process).
TN: No formal evaluation—management implements results on case-by-case basis.

QUESTIONS 1 AND 2

STATE	Q1	Q2					
	Formal measure exists?	Project selection	Project management	Post-project implementation	Staff productivity	Program benefits	Other
AK*							
AZ	Y		Y				
CA	Y	Y	Y				
CO	Y	Y	Y	In progress			
CT	Y	Y	Y	Y	Y	Y	
DE	N						
FL	Y	Y	Y (see below)	Y	Y	Y	
GA	N						
HI	N						
IL	Y			Y			
KS	Y					Y	
KY	N						
LA	Y						See below
ME	N						
MD	Y	Y	Y	Y		Y	
MN	Y	Y	Y	Y			
MS	Y				Y		
MT	Y		Y	Y			
NE	N						
NV	N						
NH	N (see below)						
NJ	N						
NY	Y		Y			Y	Y (see below)
OK	Y (begin July 2000)		Y				Y (see below)
PA	Y	Y	Y				Y (see below)
RI	N						
SC	N						
TN	N						
TX	Y			Y			Y (see below)
UT	Y	Y	Y	Y	Y	Y	Y (see below)
VT	N (see below)						
VA	N (see below)						
WY	N (see below)						

*Indicates questionnaire response not received for this questionnaire, but at least one other unit did provide response.
FL: Related to delivery of expected products, not on-time, on-budget performance.
LA: Numbers of projects started, completed, and continuing. LA DOTD is currently updating strategic plan and developing performance indicators (not measures) for the areas noted above in Question 2 (see LTRC Report, "Development of Performance Indicators for DOTD Programs," by T.G. Ray, May 1998).
NH: Has done little to date with performance measures—efforts concentrated in 2001.
NY: The number of technology transfer events is measured because of the difficulty in measuring the incremental benefit of each event. Also, the costs of providing technical consultation (problem solving) are reported, but the benefits would be reported elsewhere.
OK: Customer satisfaction.
PA: Customer satisfaction.
TX: Measure number of colleges and universities with active research projects; measure number of total active research projects.
UT: New products evaluation program.
VT: Provided responses via personal interview instead of questionnaire. No formal performance measure mechanisms in place.
VA: Provided a summary of performance indicators (not measures) recently put into place in the areas of: customer satisfaction (customer satisfaction surveys, customer anecdotes, requests for library services or publications); innovations implemented (number of specifications revised, number of recommendations implemented, number of facilities with extended service life, number of legislative or policy changes, dollar value of costs saved/avoided); advancing the state of the art (number of refereed publications, number of presentations at conferences, number of committee memberships, number of reports published, number of "hits" on website); personnel development (number of graduating transportation students supported by VTRC at Univ. of Virginia and VPI, number of courses/workshops taught by VTRC staff, number of LTAP-sponsored courses, number of LTAP workshop participants).
WY: Some information provided on WY practices for areas not formally measured.

QUESTION 3(A)—PROJECT SELECTION/PROGRAMMING

STATE				
	Qualitative measure	Quantitative measure	Written description	How measured
AK				
AZ				
CA	Y	Y	Y	Cost/benefit
CO	Y		Y	Do projects selected match strategic direction?
CT	Y	Y	Y	Per State Transportation Research Manual
DE				
FL	Y	Y	N	Distribution of funds & number of projects compared to needs and available funding
GA				
HI				
IL				
KY				
KS				
LA				
ME				
MD	Y	Y	N	
MN	Y	Y	Y	From process on Research Proposal Evaluation Form
MS				
MT				
NE				
NV				
NH				
NJ				
NY				
OK				
PA		Y	Y	Matrix used for tracking; also, customer satisfaction survey card used
RI				
SC				
TN				
TX				
UT	Y	Y	Y	Annual workshop held; diversity of participants and projects
VA				
WY				

Question 3(a)-continued

STATE			
	Measurement frequency	Why measured	How benchmarks established
AK			
AZ			
CA	Project selection cycle	Ensure that projects address DOT issue	Research Program Advisors Council reviews
CO	Annual project selection cycle	Ensure projects focus on important department issues	Research and Implementation Council develops with Chief Engineer
CT	Project selection cycle	Provides accountability to problem submitters	By synchronizing project selection process with major research program schedules for work program development
DE			
FL	Annually	Satisfy department needs	Discussions with top management
GA			
HI			
IL			
KY			
KS			
LA			
ME			
MD	Annually		
MN	Annually	To select the “right” projects	Polling MnDOT management and customers
MS			
MT			
NE			
NV			

NH			
NJ			
NY			
OK			
PA	Monthly	Make program adjustments based on customer input	Past experience and other methods
RI			
SC			
TN			
TX			
UT	Annually	Select most pressing problems and maintain balanced program	Division, upper management, and stakeholder input
VA			
WY			

Question 3(a)-continued

STATE				
	Frequency of goal updating	Actions triggered if goals not met	Actions triggered if goals exceeded	Agency satisfied with measures
AK				
AZ				
CA	Every project selection cycle	Y–advise project leader of decision and ways to improve	N	N
CO	Annually	Y	N	Y
CT	Annually	Y (see below)	Y	See below
DE				
FL	With strategic or long-range plans	N	N	Y
GA				
HI				
IL				
KY				
KS				
LA				
ME				
MD	Annually	Y	N	N–periodically review and revise
MN	As needed	Y (project not selected for funding)	Y (funded)	Continually improving
MS				
MT				
NE				
NV				
NH				
NJ				
NY				
OK				
PA	Annually	Y	N	N–looking for improvements
RI				
SC				
TN				
TX				
UT	Annually	Y	N	Y
VA				
WY				

CT: Special efforts are made to complete project selection on time for targeted research programs.
Goals are focused on maintaining a responsive project selection process.

QUESTION 3(B)—PROJECT-LEVEL MANAGEMENT

STATE				
	Qualitative measure	Quantitative measure	Written description	How measured
AK				
AZ		Y	N	Process/progress status report
CA	Y	Y	Y	Project work plan and costs
CO	Y		Y	Individual performance objectives in researcher's performance evaluations
CT	Y	Y	Y	Per State Transportation Research Manual
DE				
FL	Y		Y	Completeness of research
GA				
HI				
IL				
KY				
KS				
LA				
ME				
MD	Y	Y	Y	
MN	Y	Y	Y	Use contract expiration date as indication of timeliness. Use deliverables descriptions to evaluate content.
MS				
MT	Y		N	Technical panel judges content of project; Research Manager tracks timeliness and budget
NE				
NV				
NH				
NJ				
NY		Y	Y	Time and budget compared to original projections
OK		Y	Y	Adherence to deadlines
PA		Y	Y	Matrices of tracking information used
RI				
SC				
TN				
TX				
UT	Y	Y	Y	Through observation by technical advisory committee (TAC) and project manager for each study
VA				
WY	Y		N	Based on research contract

Question 3(b)-continued

STATE			
	Measurement frequency	Why measured	How benchmarks established
AK			
AZ	Quarterly	To track progress	
CA	Quarterly	To assist project mgrs and program mgmt	Based on each project's work plan
CO	Annually	On-time completion provides some indication of benefit from research	Project work plans
CT	Quarterly	To instigate corrective action	Through annual work program development processes
DE			
FL	Each project		Peer reviews
GA			
HI			
IL			
KY			
KS			
LA			
ME			
MD	Annually	Ensure that customer needs are met	
MN	Quarterly (goal)	Closely measured research helps achieve goals	Negotiation between researcher and technical advisor
MS			
MT	Continually	To ensure efficient receipt of desired product	Negotiation between researcher and technical panel
NE			

NV			
NH			
NJ			
NY	Annually	To determine if researchers plan their projects well and if the clients receive products when expected	Time and budget as originally projected
OK	Annually		
PA	Monthly	To track project timeliness and budget issues	Past experience
RI			
SC			
TN			
TX			
UT	Annually	To create good work plans, get appropriate TAC membership, keep projects on schedule, and product quality deliverables	TAC feedback and region/division visits
VA			
WY	Quarterly per project	Accountability, to lead to credibility and viability	

Question 3(b)-continued

STATE				
	Frequency of goal updating	Actions triggered if goals not met	Actions triggered if goals exceeded	Agency satisfied with measures
AK				
AZ				
CA	As needed	Y	Y	Y
CO	At project initiation	Y–lower individual performance rating	Y–higher individual performance rating	See below
CT	Annually (or as needed)	Y–correction action	Y–varies based on nature of performance	See State Transportation Research Manual
DE				
FL	With every strategic or long-range plan	Y	N	
GA				
HI				
IL				
KY				
KS				
LA				
ME				
MD	Annually	Y	N	N–review and update periodically
MN	On project basis (time needs sometimes change)	Y (withhold payments)	May pursue patent	Unsatisfactory timeliness suggests improvement needed
MS				
MT	As needed (on project basis)	Y (PI is contacted)	N	Moving toward more objective measures
NE				
NV				
NH				
NJ				
NY	When updating the Policy and Procedural Manual	Y (but not helpful)	N	Most projects are completed near budget but over schedule, which is problematic if the schedule problems result from poor planning (instead of actual research problems)
OK	Annually	N	N	
PA	Annually	Y	N	Y
RI				
SC				
TN				
TX				
UT	Annually and at conclusion of project	Y	N	N
VA				
WY		Y (not formal/written)	Y (not formal/written)	Y

CO: Changed conditions and poor performance from university researchers complicate process.

QUESTION 3(C)—POST-PROJECT IMPLEMENTATION

STATE				
	Qualitative measure	Quantitative measure	Written description	How measured
AK				
AZ				
CA			N	Measure is under development
CO				
CT	Y	Y	Y	See StateTransportation Research Manual
DE				
FL	Y	Y	N	Actual vs. planned activities
GA				
HI				
IL	Y	Y	Y	
KY				
KS				
LA				
ME				
MD	Y	Y	Y	
MN	Y	Y	Y	Close-out memo discussing implementation is initiated
MS				
MT	Y		Y	
NE				
NV				
NH				
NJ				
NY				
OK				
PA	Being developed			
RI				
SC				
TN				
TX		Y	Y	Percent and number of research recommendations implemented within 2 years of project completion
UT	Y	Y	Y	Feedback on completed studies at quarterly implementation meeting. Track for 2+ years.
VA				
WY				

Question 3(c)-continued

STATE			
	Measurement frequency	Why measured	How benchmarks established
AK			
AZ			
CA		Unused research results wastes money	
CO			
CT	Annually	Document value of research to ConnDOT	Varies by nature of project
DE			
FL	Project selection cycle		Research Center Manager develops
GA			
HI			
IL	About 3 yrs after project completion	Goal of research program is to change current practice	
KY			
KS			
LA			
ME			
MD	Annually	Determine usefulness of research	Discussions between researcher and user
MN	Project-specific (tracking system)	Determine return on investment	Identified in implementation planning process
MS			

MT	On completion of project	Implementation is considered essential indicator of quality of program	
NE			
NV			
NH			
NJ			
NY			
OK			
PA			
RI			
SC			
TN			
TX	Quarterly estimate; annual determination	Determine if research results are implementable: determine how successful TxDOT is in implementing new products and processes	By state legislature
UT	Quarterly	Evaluate the quality of deliverables over time; put resources to implementation activities; select T^2 strategies	Action items are assigned at quarterly meeting
VA			
WY			

Question 3(c)-continued

STATE				
	Frequency of goal updating	Actions triggered if goals not met	Actions triggered if goals exceeded	Agency satisfied with measures
AK				
AZ				
CA		N	N	N
CO				
CT		Y–Manager takes corrective action	Y	Y–see State Transportation Research Manual
DE				
FL	With every strategic or long-range plane	N	N	
GA				
HI				
IL				N (see below)
KY				
KS				
LA				
ME				
MD	Annually	Y	N	N–review and revise periodically
MN	Not updated (once established for a project)			Y
MS				
MT	As needed	N	N	N–moving toward more objective approaches
NE				
NV				
NH				
NJ				
NY				
OK				
PA				
RI				
SC				
TN				
TX	Every 2 years	N	N	See below
UT	Annually	Y	N	N–need new tools to monitor implementation and T^2
VA				
WY				

IL: This measure does not reflect the importance of research that shows something should not be implemented. The measure is also insensitive to the "problem solving" nature of research that is important to top management.
TX: Trying to change from "percentage of researcher recommendations implemented" to "percentage of products resulting from research implemented."

QUESTION 3(D)—STAFF PRODUCTIVITY

STATE				
	Qualitative measure	Quantitative measure	Written description	How measured
AK				
AZ				
CA				No good method developed—extremely difficult
CO	Y			Individual performance plans and evaluations
CT	Y	Y	Y	See State Transportation Research Manual
DE				
FL	Y		Y	Review & performance planning
GA				
HI				
IL				
KY				
KS				
LA				
ME				
MD	Y			
MN	(in development)			
MS	Y	Y	Y	Agency-wide performance evaluation criteria
MT				
NE				
NV				
NH				
NJ				
NY				
OK				
PA				
RI				
SC				
TN				
TX				
UT	Y	Y	N	Staff plans are tied to department goals. Track the number of projects managed vs. planned and measure expenditures vs. planned
VA				
WY	Y	Y	N	Subjective estimation by questioning end users

Question 3(d)-continued

STATE			
	Measurement frequency	Why measured	How benchmarks established
AK			
AZ			
CA			
CO	Annually	State law requires	Negotiation between employee and supervisor
CT	Quarterly	Control progress of work	Through annual work program
DE			
FL	Annually		Research Center Director develops
GA			
HI			
IL			
KY			
KS			
LA			
ME			
MD	Annually	All employees participate in annual performance review	Through discussions (supervisor and employee)

MS	Annually	Ensure staff performance	Supervisors decide
MT			
NE			
NV			
NH			
NJ			
NY			
OK			
PA			
RI			
SC			
TN			
TX			
UT	Annually	Feedback into process, such as "Is training needed for staff?"	Educated guess
VA			
WY	Occasionally	To show how research is paying off	None

Question 3(d)-continued

STATE				
	Frequency of goal updating	Actions triggered if goals not met	Actions triggered if goals exceeded	Agency satisfied with measures
AK				
AZ				
CA				
CO	Annually	Y–low rating	Y–high rating (see comment)	See below
CT	Annually	Y	Y	Y–system is acceptable
DE				
FL	Annually	Y	Y	
GA				
HI				
IL				
KY				
KS				
LA				
ME				
MD	Annually	Y	Y	Annual review and revision as needed
MN				(measures in development)
MS	As needed	Y	N	Y
MT				
NE				
NV				
NH				
NJ				
NY				
OK				
PA				
RI				
SC				
TN				
TX				
UT	Annually	Y	Y	N
VA				
WY				N

CO: Someday exceeding or not meeting goals may be related to pay.
Unsatisfied with process because so much judgment is involved and it is perceived as not adjusting adequately for the different requirements of different jobs

Question 3(E)—PROGRAM-LEVEL BENEFITS

STATE				
	Qualitative measure	Quantitative measure	Written description	How measured
AK				
AZ				
CA	Y	Y	N	Estimated cost/benefit
CO				
CT	Y	Y		See below
DE				
FL	Y	Y	Y	See below
GA				
HI				
IL				
KY				
KS		Y	Y	Total benefits divided by total program cost
LA				
ME				
MD		Y	N	Meeting of goals and objectives
MN	(measures in development)			
MS				
MT				
NE				
NV				
NH				
NJ				
NY		Y	Y	Overall benefits (capital/operational savings, etc.) for completed projects are determined and compared to total research program costs
OK				
PA				
RI				
SC				
TN				
TX				
UT	Y	Y	Y	Have programs graded by end users. Attempts made to estimate monetary benefits.
VA				
WY	Y	Y	N	By response of end users

CT: Measures specific to a program are used.
FL: Office of Policy Planning and Governmental Accountability Transportation Commission—Performance Measures used.

Question 3(e)-continued

STATE			
	Measurement frequency	Why measured	How benchmarks established
AK			
AZ			
CA	Irregular- based on political requests		Program-level C/B, not project-level
CO			
CT	Quarterly	Document benefits	In annual work program
DE			
FL	Annually	Mission support	Not known
GA			
HI			
IL			
KS	Annually	Upper management decision	No benchmarks established
KY			
LA			
ME			
MD	Annually		Mutual agreement
MN			

MS			
MT			
NE			
NV			
NH			
NJ			
NY	Annually (but benefit-cost ratio is three-year average to smooth out spikes)	To evaluate the benefit/cost ratio of the program	Target B/C ratio is set to 1.0 (see below)
OK			
PA			
RI			
SC			
TN			
TX			
UT	Annually	Shows value of research; identifies the types of studies that are successful; improves the process	Staff input with management approval
VA			
WY	Occasionally		

NY: Benefit/cost ratio of 1 seems low, but other services the program provides increase the overall value of the program to the department.

Question 3(e)-continued

STATE				
	Frequency of goal updating	Actions triggered if goals not met	Actions triggered if goals exceeded	Agency satisfied with measures
AK				
AZ				
CA	Periodic, irregular	N	N	N
CT	Annually	Y	Y	Y
DE				
FL	Annually	Y	N	Y
GA				
HI				
IL				
KS				Y (see below)
KY				
LA				
ME				
MD	Annually	Y	Y	Review and revised as needed
MN				(measures in development)
MS				
MT				
NE				
NV				
NH				
NJ				
NY	When updating Policy and Procedure Manual	Y (management attention alerted)	N	Y (use of $$$ B/C ratio easily understood by upper management)
OK				
PA				
RI				
SC				
TN				
TX				
UT	With strategic or long-term plans	N	N	N
VA				
WY				

KS: Obtaining accurate and complete benefit-cost ratio information is a continuing challenge.

QUESTION 3(F)—OTHER

STATE				
	Quantitative measure	Qualitative measure	Written description	How measured
AK				
AZ				
CA				
CT				
DE				
FL				
GA				
HI				
IL				
KY				
KS				
LA				
ME				
MD				
MN				
MS				
MT				
NE				
NV				
NH				
NJ				
NY	Y	Y		Number of T^2 activities measured by staff participation and number of events; Cost of consultation efforts is reported
OK	Y		Y	Customers surveyed as projects are completed
PA	Y	Y		
RI				
SC				
TN				
TX		Y	Y	By simple tally of projects and universities under contract
UT	Y	Y	Y	Percent of products evaluated; percent of complaints resolved; value added
VA				
WY				

Question 3(f)-continued

STATE			
	Measurement frequency	Why measured	How benchmarks established
AK			
AZ			
CA			
CT			
DE			
FL			
GA			
HI			
IL			
KY			
KS			
LA			
ME			
MD			
MN			
MS			
MT			
NE			
NV			
NH			
NJ			

NY	Annually	See below	No benchmarks
OK	Annually		
PA	As needed (3-4x/year)	Determine if program is helpful to customers	Customer Service Index guidelines
RI			
SC			
TN			
TX	Quarterly	Measures size of program and diversity of program participants	State legislature
UT	Annually	Improve the introduction of new products and processes	Staff input with management approval
VA			
WY			

NY: T^2 activities measured to report on type and number of activities; consultation costs to report involvement in solving clients' problems.

Question 3(f)-continued

STATE				
	Frequency of goal updating	Actions triggered if goals not met	Actions triggered if goals exceeded	Agency satisfied with measures
AK				
AZ				
CA				
CT				
DE				
FL				
GA				
HI				
IL				
KY				
KS				
LA				
ME				
MD				
MN				
MS				
MT				
NE				
NV				
NH				
NJ				
NY				T^2–No (searching for method that does better than just represent a presence of the activity) Consultation–Yes
OK	Annually	N	N	
PA	Each project	Y	N	Y
RI				
SC				
TN				
TX	Every 2 years	N	N	Y
UT	Every 2 years	Y	N	N–currently a Quality Improvement Team is reinventing this process
VA				
WY				

STATE DOT RD&T PROGRAMS
QUESTION 4

(Primary reason(s) shown with ^; if no specifics noted, an X is shown)

STATE						
	Project selection	Project mgmt	Post-project implementation	Staff productivity	Program benefits	Other
AK						
AZ	B, E^		E^	B, E^		
CA			C^, D	B,C,F^		
CO			X		X	
CT						
DE		C^, D, E	C^, E		B, C^, E	
FL						
GA			C^, D		B, C^	
HI	B, C, E^	B, C, E^	B, C, E^	B, C, E^	B, C, E^	
IL	X			X		
KS						
KY	A^		E^	D^, E		
LA		E (see below)	E	E	E	
ME			E, F^ (see below)		C, E^	
MD			C^			
MN						
MS						
MT	E, F^			E, F^	E, F^	F (see below) inadequate staffing
NE			B^		B^	
NV	B^	E^	E^		B^	
NH		X	X	X	X	
NJ	F^ (see below)	F^	F^	F^	F^	F^
NY	A^		D^	A^, B, D		
OK	B, C^		B, C^		B, C^, E	
PA				A, B^, D	D^	
RI			C^, E		B^, C, E	
SC	A, B^, E		B^, E		B^, C, E	
TN		B^	B, E	B^	B^	
TX		B, E^			B, C, E^	
UT						
VA						
WY			C^, E		B, C^, E	

CA – F^: Results can be skewed by addressing what is measured, not the overall usefulness to the department.
LA: For all—measures not useful, indicators may be.
ME: Because implementation of research results is in the hands of other units and is not institutional, it is difficult to use as a performance measure for the research unit.
MI: inadequate staffing.
NJ: Performance measurements will be done during the third year of current strategic plan (currently in Year 1—building the organization).

STATE DOT UPPER MANAGEMENT

STATE					
	Familiar with RD&T performance measurement	Use same or different method to evaluate RD&T	Perceived effectiveness	Critical information	Non-RD&T units using PMs
AK	No formal PMs exist	Timeliness & budget adherence	4	Ratio of research costs to design/construction/maintenance savings	Design Construction
AZ					
CA	N	N–Use specific projects or emphasis areas.	7	Departmental savings; reduction in system delays and accidents	Construction Maintenance Project Development
CO	Y	N–actual application of research conclusions is tracked	6	Conclusions of the research (good or bad) and action plan for implementation	See below
CT	(Construction Admin) Y (Chief Engineer) Y	Y Y (see below)	8 8	Resources and time required to implement project results; projected and actual benefits from implementation Quantifiable information on research implementation results	Construction Maintenance Engineering Design Right-of-Way Public Transportation (bus, rail, aviation, ferries, ports)
DE					
FL					
GA	N	N	7	Documentation of results of research initiatives that have been incorporated into business practices	Design, Maintenance Construction, Planning
HI	Y	Y	8	Effectiveness of results	All units
IL	Y	Y	8	Implementable products	Y (see below)
KS	Y	Y	8	Number of implementable products and cost-benefit	Y (noted in strategic plan)
KY					
LA	Y	N	6	Practical application of results	Project letting Maintenance
ME	(Planning, Research, & Community Services) N (Maintenance & Operations) Y	 Y	5 7	Measures of efficiency (cost-effectiveness, productivity) and improved performance (quality, service life) Implementation of MDOT or other DOT research that results in improvement	Planning
MD	(Chief Engineer – Ops) N (Planning Director) Y	N Y	6 5	Improved operational efficiency (e.g., safety and system performance) Documentation of implementation and benefits	Y (noted in Business Plan) All offices (SHA Business Plan)
MN					
MS	Y	Y	7	Implementation of findings and identification of problems for future research	Y (agency-wide for program-level staff productivity)
MT	Y	N	6	Amount of research implemented by MDT	In progress
NE	Y	N	8	Types of infrastructure with lower maintenance; cost-effectiveness of new facilities/procedures; before-and-after safety statistics	Y (agency-wide document)
NV					
NH	See below				
NJ	(Capital Program) N	N	5	Number of projects implemented and time to implementation	Design Construction

NJ	(Operations) N (Planning, Research, & Local Gov't) Y	N Y	9 9	Success in completing individual projects that can impact department operations	Recently initiated several Y
NY	Y	N	8	Benefit-cost ratio; documentation that RD&T is solving problems of interest to the program areas it serves	Construction Design Maintenance
OK	(Deputy Director) N (Chief Engineer) N (Asst. Director – Operations) Y (Asst. Director – Preconstruction) N	 N Y N	 3 7 5	 Timely completion of research on issues; projects selected as being of interest to ODTO Return on investment Return on investment; useful results over 10 yr, 5 yr, 2 yr periods	ODOT implementation of performance measures established in 2000 for the first time N Maintenance Just starting
PA					
RI					
SC	Y	Y	6	Cost vs. savings	N
TN	N	Y	6	Percentage of projects implemented	Y (strategic-plan related)
TX	Y	Y	9	Focus on implementation of results/products and cost savings	Y
UT	Y	Y	8	See below	Y
VA					
WY					

CA, CO: Has unspecified benchmarks, but no actions are triggered if goals unmet.
CO: Quantifiable measures for all units under development.
CT: Performance is tracked within work scopes, budgets, and time frames; remedial measures instituted if goals not met.
IL: Changing department-wide PMs to use the "balanced scorecard" approach by Kaplan & Norton.
NH: Has done little with performance measures agency-wide; results reported by RD&T unit.
UT: Summary information indicating new products or technologies that have been successfully implemented and that have measurably resulted in one or more of the following: increased efficiency, reduced costs, improved quality, reduced maintenance, extended product life cycles.

STATE DOT RD&T CUSTOMER RESPONSES

STATE	1	2	2	3
	Customer unit(s)	Familiar with RD&T performance measurement	Use same or different method to evaluate RD&T	Perceived effectiveness
AK*				
AZ	ITD	N	N–Does the outcome answer the questions?	8
	Transportation Technology Group	N	Y–Results in changes to ADOT specs and processes	7
	Holbrook District	Y	Y	7
	Bridge	N		5
CA				
CO	Bridge	N–only anecdotally	N	8
CT	Fiscal Admin	N	Y (see below)	6
	District Maintenance	Y	Y	8
	Pavement Mgmt	N		7
	Bridge Safety	N	Y	5
DE				
FL	Environmental Mgmt Office	Y	Y	9
	Materials	Y	Y	7
	Scheduling	N		9
GA	Maintenance	Y	Y	8
	Bridge	N	N	3
	Planning	Y	N	8
HI	Oahu District	N	N	8
	Materials (several responses)	N	N	~8
IL	District 6	N	N	10
	District 8	Y	N	
	Anonymous unit	Y	Y	9
KS				
KY				
LA	Operations	Y	Y	8
ME	Management Systems	N	N	8
	Bridge Management	N		6
	Testing	N		8
	Freight Transport	Y	N	8
MD	Construction	Y	Y	6
	Materials Technology	Y	Y	5
	Traffic & Safety	Y	Y	8
	ITS (CHART)	Y	Y	9
	Maintenance	Y		7
	Planning & Preliminary Engineering	Y	Y	7
MN	Environmental Services	Y	Y	8
	Maintenance	Y	Y	7
	Metropolitan Division	N	N	7
	Bridges & Structures	Y	Y	9
	Materials	Y	Y	8
	Freight, Railroads, & Waterways	Y	N	6
	Traffic Engg	N		6
	Standards			8
	Construction			6
MS				
MO	Materials	N	N	5
MT	Environmental Services	N	N	8
NE	Roadway Design	N		7
	Bridge	N	N	8
NV	Materials	Y	N	8
	Bridge	N	N	8
NH				
NJ	Traffic Operations	Y	N	7
	Civil Engineering	N	Y	7
	Structures	N	N	8
	Design Services	N		5
	Operations Support	N	N	8

NJ	Materials	N		5
NY				
OK	Division Engineer	N	N	2
	Roadway Design	Y	Y	7
	Maintenance	N	N	4
	Planning	N	Y	6
	Division Engineer	N	Y	7
	Division Engineer	Y	Y	7
	Traffic			8
PA	Various units (see below)	N	N	7
		Y	Y	7
		N		8
		N		8
		N	N	8
		N	N	5
		N	N	2
		N	N	6
		Y	Y	6
		Y	N	8
		N	N	8
		N	N	6
RI				
SC	Maintenance	N		8
	Traffic	Y	Y	8
	Bridge Design	N	N	9
TN				
TX				
UT	Engineering Services	Y	N	6
	Anonymous unit	N	N	4
	Region Director	N	N	7
	Traffic Operations	N	N	2
VA				
WY	Bridge	N	N	8
	Geology	N	N	10
	Anonymous unit	N		4
	Wetlands and Wildlife	N		8
	Materials	N	Y	8

PA: Operating units provided responses on a mistaken understanding of some anonymity; therefore, units are not identified.

STATE*	4	5**	5**	6
	Contacted for feedback	Feedback on "payoff"	Biggest need from RD&T	Operating unit PM
AK				
AZ	Routinely	Occasionally	Continually verify material performance	Y
	Rarely	Occasionally		Y
	Occasionally	Occasionally	How do state-of-the-art technologies really affect construction and maintenance?	Y
	Occasionally	Occasionally	Continue projects on drilled shaft design and construction	Y
CA				
CO	Routinely	Routinely	Continue current fine level of service	Y
CT	Occasionally	Occasionally	Specific project on videologging	Y–on-time completion of tasks
	Rarely	Rarely	Provides reports to track projects; more info on SHRP	Y–miles of pavement rated or number of projects completed
	Occasionally	Occasionally	Timely circulation of info; continue informational workshops	N
	Rarely		Discussions on future needs of bridge safety unit	N
DE				
FL	Routinely	Routinely	More funding for environmental issues	Y–short- and long-term goals
	Rarely	Routinely	Well-defined strategic goals; website for information; funding for high priority projects	N
	Routinely	Routinely	Finance projects to address specific concerns	

GA	Occasionally	Occasionally	More interaction with RD&T on "real projects" to incorporate findings	Y
	Rarely	Occasionally	Provide support with research in what we are doing and what we need	Y (somewhat)
	Routinely	Routinely	Continued openness and flexibility regarding new research projects	Y (time-based)
HI	Routinely	Routinely	Provide useful tools to evaluate unit activities and overcome deficiencies with training	N
	Rarely	Occasionally	Keep funding needed research	N
IL	Rarely	Rarely	No District 6 and RD&T interface currently	Y (in development)
	Routinely	Routinely	Continue central clearinghouse role for new products/processes to avoid district-by-district evaluations	Y (qualitative)
	Routinely	Occasionally	Response to new products and implementation	N
KS				
KY				
LA	Occasionally	Occasionally	Technical assistance; short turnaround tasks	Y (established by legislature)
ME	Routinely	Occasionally	More marketing of research products and efforts	Y (not comprehensive)
	Rarely	Rarely		N
	Routinely	Occasionally	Investigates new products/technology to determine where to implement	Y (cost-based)
	Occasionally	Occasionally	Internet searches on emerging subject areas	Y
MD	Occasionally	Occasionally	Provide more info on new technology as it relates to construction	Y
	Rarely	Occasionally	Support of program goals	Some
	Occasionally	Occasionally	Provide funds; keep (units) informed; coordinate overall RD&T programs	Some
	Occasionally	Occasionally	Support university research program; identify funding for ITS research; keep ITS unit informed of national research programs and findings	Y
	Routinely	Routinely	Keeping unit abreast of new technologies	Y
	Occasionally	Occasionally	Continue proactive coordination of the researcher with the requests made by the planning/engineering office	Y (in draft form)
MN	Routinely	Routinely	Provides a means to answer some difficult questions	N
	Routinely	Rarely	Help stay abreast of current trends and technology/information	Y (not regularly evaluated)
	Routinely	Routinely	Continue to find ways to fund RD&T applications for operational evaluation	Being developed
	Routinely	Routinely	Strive for effective administrative liaison role to ensure that problems become projects that become solutions	Y
	Routinely	Routinely	Provide useful products	Y
	Occasionally	Occasionally	Locate researchers who are willing to answer questions/needs (too often, the research compromises to meet the researchers' interests)	Y
	Occasionally	Occasionally	Provides an avenue for sharing ideas and common themes	Y
	–	–	Provide funding to do more research	
	Occasionally	Routinely	There is not much research related directly to contract and construction management	Y
MS				
MO	Occasionally	Routinely	Study product results of pilot projects using non-specification technologies/items	Y

MT	Routinely	Occasionally	Keep unit aware of latest research; continue to ask for input and participation	N
NE	Occasionally Occasionally	Occasionally Occasionally		N Y
NV	Routinely Routinely	Routinely Occasionally	Research unit does the administration so that bridge staff can (focus on and) coordinate the technical issues	N N
NH				
NJ	Occasionally Routinely Routinely Don't know Routinely Occasionally	Occasionally Routinely Routinely Don't know Occasionally Occasionally	RD&T unit addresses mid-term and long-term problems that traffic unit faces CADD efficiency improvements Facilitates contact w/ research investigators and acts as point of contact on research efforts Continue to ask for input on what research to perform Identify more cost-efficient products/services Ensure that implementation plan is part of the research	Partial (tracking of goals and results) Y Y Y Y N
NY				
OK	Rarely Occasionally Rarely Occasionally Occasionally Routinely Routinely	 Occasionally Occasionally Occasionally Occasionally Occasionally	Make reports more practical and less technical Provide analytical expertise; provide and facilitate implementation plans Provide guidance to the best literature (esp. economic impacts) Advancing our knowledge to improve design, construction, and maintenance Provide information on products/methods performed by maintenance crews which do not have designer support More data-processing support	N N N Y Being developed N N
PA	Occasionally Occasionally Routinely Occasionally Routinely Rarely Rarely Occasionally Occasionally Routinely Routinely Occasionally	Occasionally Occasionally Occasionally Rarely Routinely Rarely Occasionally Routinely Occasionally	Provides administration so that project manager can focus on technical tasks Track implementation Work with division contact Provides administration so that project manager can focus on technical tasks They help projects to happen now Speed up approval process Shorten project selection process to shorten overall research time Quicken work order approval Fast implementation Serves as facilitator Helps developing research needs	(Varies)
RI				
SC	Occasionally Routinely Routinely	Occasionally Routinely Occasionally	Continue communication within department to determine needed research or advise of ongoing research Continue valuable work Good contact with local universities for small research projects	N Work compared to goals in strategic plan Adherence to strategic plan and number of projects to complete
TN				
TX				
UT	Occasionally	Occasionally	Evaluate the approved product list so that it will be compatible with	Y

			standard specifications	
	Occasionally	Rarely	Proof of concept projects are a waste of time; need practical products	Y
	Occasionally	Occasionally	Keep up good job of making sure the research topics are field generated	Y
	Occasionally	Rarely	Screen and thoroughly test ITS products	Y
VA				
WY	Occasionally	Occasionally	Keep unit abreast of ongoing research and issue periodic call for proposals	Y
	Routinely	Occasionally	Keep up the library and search database for information	N
	Rarely	Occasionally	RD&T funds problems that do not fit into traditional areas (sometimes) – for example, GIS was initially funded by Research. RD&T should be geared toward more actual research, but many times we do not have good candidate projects.	N
	Occasionally	Occasionally	Provide research $$$ and help prepare proposals	
	Occasionally	Occasionally	Support research and implementation activities as requested by (our unit)	N

*Indicates that questionnaire response not received from customer unit(s), but response was received from RD&T or management unit.
**Questionnaire form had typographical error noting two occurrences of "Question 5."
CA: Developing its own customer "payoff" feedback process currently.
CT: Fiscal administration assumes that research measures its performance based on completion of task in timely manner.

APPENDIX C

New York, Utah, and Virginia Summaries of Performance Measure Structure

TRANSPORTATION RESEARCH & DEVELOPMENT BUREAU—DIRECTOR (New York)

Function	Performance Measure	Goal	Used				Benefits/ Comments
			How	When	By Whom	For What	
Conduct Engineering Research Projects (both in-house consultation and ERTAP)	1) Project narratives in monthly report 2) Semi-Annual Progress and Expenditure Report (SPR) Part II 3) Maximize benefit/cost ratios while satisfying client	1) Research is progressing in a timely manner and satisfies client 2) Keep research program under budget and deliverables arriving on schedule 3) Research is conducted in a cost-effective manner	1) To monitor the progress of each project and to ensure that client is satisfied 2) For each project, semi-annually check expenditures and see that progress is being maintained according to original work plan 3) To see that research is cost-effective to Department	1) Narratives for each project are included in Monthly Report(s) 2) Used semi-annually (based on Federal Fiscal Year) 3) Done at completion of each project	Bureau director FHWA Division director (if noteworthy) Section heads	Evaluation of ongoing research program Prioritize resource allocation Planning research of future program Facilitate FHWA oversight	Both performance measures ensure that research will provide *direct* financial benefits to Department (applied as opposed to theoretical research)
Administration/ Publications	Newsletters & Reports produced (included in Monthly Report)	Reports, newsletters, etc., published in a timely manner and conforms to quality standard	To document reports/newsletters that have been published and to whom they have been distributed Document other admin./ publ. activities	Completed reports and their distribution are included in the Monthly Report	Bureau director Division director (if noteworthy) Section heads	To ensure that rpts./publ. are of consistent quality and are equitably distributed within Department	All reports are edited according to TRB standards
Library Functions	Computerized log of all library transactions	The Bureau has necessary research material to perform its functions	To keep track of all library transactions	Library activities are noted in Monthly Report Transactions log tabulations are also documented semi-annually in SPR Part II	Admin. Asst. Librarian Bureau director	Accurate record keeping Planning of future operations	Library is utilized by many other bureaus in Department
Technology Transfer/ Training	Narratives of tech. transfer activities in Monthly Report Monthly Report narratives of training activities and course evaluations	Implement new technology throughout entire Department Provide training as requested and client agencies are satisfied	To keep track of tech. transfer activities Make note of all training activities and document	Tech. transfer/ training activities are narrated in Monthly Report Course evaluations are carried out at completion of the training session(s)	Bureau director Division director (if noteworthy) Section heads Associate statistician (for statistical training)	Documentation of technology transfer and training activities To see that training was performed to clients satisfaction	Ensures statewide implementation of research and new tech. developments Bureau performs statistical consultations for entire Department

TRANSPORTATION RESEARCH & DEVELOPMENT BUREAU—ADMINISTRATION/PUBLICATIONS (New York)

Function	Performance Measure	Goal	Used				Benefits/ Comments
			How	When	By Whom	For What	
Coordinate Fiscal Management of Departments Semi-Annual Progress and Expenditure Report of SPR Part II Research Program	Records are kept up to date	Accurately update records in a timely, detailed fashion	To monitor that research program is running under budget and on time	For publication of Annual Work Plan SPR Part II (Oct. 1) Semi-Annual Work Plan (Mar. 30)	FHWA Bureau director Section heads Project supervisors Division director Planners Regional Personnel	To collect and coordinate all research information into one source	Facilitate FHWA oversight Systematically documents and monitors current research program
General Admin., including: Budget Personnel Purchasing Travel Equipment Mail	All administrative records, forms, and procedures are completed timely and accurately	Work is completed on time Errors are minimized Stay responsive to needs of the Bureau	Used to keep Bureau running smoothly, effectively, and financially sound	Continuously	Division director Bureau director Admin. asst. Relevant bureau personnel	Justify resource/ personnel decisions	Bureau is running smoothly and fiscally responsible Bureau is forced to plan future activities
Library	Timely collection, retrieval, and distribution of pertinent research material	Computerized operations Establish budget for enlarged collection Write new mission statement	To monitor distribution of materials and services provided	Continuously	Division director Bureau director Admin. asst. Librarian	To track library activities and plan future acquisitions	Library efficiently provides required services
Publications	The timeliness, quality, and standards/ format conformance of all edited research publications (i.e., reports and newsletters)	Publications are edited within reasonable time expectations and according to accepted tech. and editorial standards Develop editorial "format" standards for all Bureau publications	To monitor reports, track their distribution, and develop editorial standards	Continuously, as milestones activities (i.e., project completion dates) arrive	Division director Bureau director Admin. asst. Editor	To monitor publications, distribution, ensure consistent style/ quality of all Bureau reports/ correspondence	Reports and newsletters published on timely basis, in accordance with appropriate tech. standards (i.e., TRB, Univ. of Chicago, etc.) Working to establish a computerized tracking and distribution system

TRANSPORTATION RESEARCH & DEVELOPMENT BUREAU—MATERIALS/PAVEMENTS (New York)

Function	Performance Measure	Goal	Used				Benefits/ Comments
			How	When	By Whom	For What	
Materials/ Pavements Related Research Projects	1) Daily contact with project engineers 2) Engineering Research Semi-Annual Progress and Expenditure Report (SPR) 3) Update for each project in Monthly Report 4) Benefit/ cost analysis	1) See that project is progressing smoothly 2) Project is on time and under budget 3) Satisfactory progress for each project 4) Benefit/cost significantly higher than 1	To see if research program is being conducted in timely and cost-effective manner	1) Daily 2) Every 6 months 3) Monthly 4) After the completion of each	Project supervisors (ERS Is) Section head (ERS II) Bureau director Division director	Used in making budgeting, personnel, and work priority decisions	Narratives of all projects are included in Monthly Reports, ensuring the status of every project is updated every month
Technical Assistance and Consultation for the Department	1) Daily contact with project engineers 2) Engineering Research Semi-Annual Progress and Expenditure Report (SPR) 3) Update for each project in monthly report 4) Benefit/ cost analysis	1) See that projects are progressing smoothly 2) Project is on time and under budget 3) Satisfactory progress for each project 4) Benfit/cost ratio significantly higher than 1	To see if research program is being conducted in timely and cost-effective manner Ensure that research can/will be used by the Department	1) Daily 2) Every 6 months 3) Monthly 4) After the completion of the Project	Section head Bureau director Division director	Used in making budgeting, personnel, and work priority decisions	Ensures that materials/ pavements related research can be used by the Department
Serve on Committees, Technical Advisory Panels, etc.	Activities are noted in Monthly Report	Awareness and networking in overall materials/pavements research program	Section Monthly Report compiled from individual monthly reports	At the end of the month	Section head Bureau director Division director (if noteworthy)	To see that state needs are addressed and state is not duplicating other research	Ensures NYS's has significant contributions to overall research community The interests of NYS are addressed by oveall research community

TRANSPORTATION RESEARCH & DEVELOPMENT BUREAU—STRUCTURES (New York)

Function	Performance Measure	Goal	Used: How	Used: When	Used: By Whom	Used: For What	Benefits/ Comments
Review Research Suggestions (for own program, contact research, etc.)	1) Necessary deadline requirements are met 2) Literature search is done and clients are properly identified 3) Clients needs are satisfied 4) Benefit/cost analysis is done 5) No. of people trained	1) Meet the program development deadlines 2) State research is not used to perform research that is already being done elsewhere; research will be used by the Department 3) Research will be used by client 4) Benefit/cost ratio is as high as possible 5) Train people as requested to implement research projects	1) To ensure timely development of the program 2) Make sure that we are not "reinventing the wheel" 3) Make sure clients are getting what they want 4) To choose the best projects to include in the program 5) To document training activities and accomplishments	1) During development of program 2) As incoming research suggestions come in and before start of project 3) Client meetings are scheduled periodically 4) In developmental stage of program 5) Included in Monthly Report	Section head Bureau director	1–4) To develop the research program to choose project that will provide the highest benefits to the state of New York and the structures research community	All research is original Research will be implemented for Department use
Conduct Research Relative to Structures	1) On time and under budget 2) Literature search is performed, client is identified 3) Client(s) is satisfied 4) No. of reports generated 5) No. of papers/ presentations 6) No. of people trained 7) Benefit/cost analysis	1) Research is conducted in a timely and cost-effective manner 2) Research is original and will have a use in structural applications 3) Document client's satisfaction 4–6) Done on demand in a timely manner 7) Benefit/cost ratios as high as possible	1) To ensure research is progressing on schedule 2) To see that research is original and there is practical usage for it 3) To show that research is or will be useful to the Department 4–6) To document work completed 7) To show that research is cost-effective	1) Semi-Annual Progress and Expenditure Report (SPR) and annual review 2) Beginning of project 3) Project's completion 4–6) Included in Monthly Report 7) At completion	Section head Bureau director Division director (if noteworthy)	Optimize research potential Program development Personnel and budget decisions	All research is original Peer review ensures quality work

TRANSPORTATION RESEARCH & DEVELOPMENT BUREAU—STRUCTURES (New York)

Function	Performance Measure	Goal	Used				Benefits/ Comments
			How	When	By Whom	For What	
Technology Transfer Activities	1) On time and under budget 2) Client satisfaction 3) No. of reports 4) Benefit/cost analysis 5) No. of people trained	1) The research is conducted in a timely and cost-effective manner 2) Client's satisfaction is determined 3&5) Done on demand 4) Benefit/cost ratios as high as possible 5) On demand	1) To ensure research is progressing on schedule 2) To show that research will have practical application 3&5) Document the work completed 4) To show that research is cost-effective	1) Show that project was conducted in a timely and cost-effective manner 2) After completion of the project, solicit client for comments on project usefulness 3&5) Included in the Monthly Report 4) Completion of project	Section head Bureau director Division director (if noteworthy)	Optimize research potential Program development Personnel and budget decisions	Ensures timely, practical, and cost-effective implementation of new technology

TRANSPORTATION RESEARCH & DEVELOPMENT BUREAU—TECHNOLOGY TRANSFER (New York)

Function	Performance Measure	Goal	Used				Benefits/ Comments
			How	When	By Whom	For What	
Conduct of Research and Technical Assistance Projects	1) Semi-Annual Progress Expenditure Report (SPR) 2) Monthly Report narratives 3) Daily/weekly contact with project supervisor 4) Project review meeting 5) Publications and presentations 6) Benefit/cost analysis	1) Under budget and on time 2) Meets needs of clients 3) See that project is progressing as intended 4) Monitor direction and progress of project and assure clients needs are met 5) Peer review is positive 6) Benefit/cost ratio is as high as reasonably possible	1) Project is completed on time and under budget 2) Update on projects' progress 3) Update on progress and direction of project 4) Presentation of the project's progress to date by principal investigator to FHWA and clients 5) To ensure quality of work 6) Resource/ personnel justification; program development	1) Every 6 months 2) At the end of every month 3) Daily/ weekly 4) Annually 5) Activities reported at end of month 6) Benefit/ cost analysis done upon completion of project	1) Project supervisors Section heads Bureau director FHWA 2) Section heads Bureau director Division director 3) Section head 4) Section head Bureau director Client FHWA 5) Section head Bureau director 6) ERS I's Section heads Bureau director	1) Communicate progress with FHWA, keep project on schedule and under budget 2&3) Keep track of each project's progress 4) Make sure direction of project is on track 5) Peer review/ quality assessment 6) Cost-effectiveness assessment	Each project's spending is monitored, making it easier to control budget Projects are reviewed for both quality and cost-effectiveness

TRANSPORTATION RESEARCH & DEVELOPMENT BUREAU—TECHNOLOGY TRANSFER (New York)

Function	Performance Measure	Goal	Used: How	Used: When	Used: By Whom	Used: For What	Benefits/ Comments
Technology Transfer Activities: 1) Distribution of TT Materials 2) Newsletters (TNT & ITS)	1) Log of what materials are sent to whom and when 2) Published on time Timely and useful articles Updated mailing lists	Increase Department's awareness of research Everybody in Department receives appropriate materials Keep Department informed	1) Keep track of activities 2) Ensure appropriate distribution of materials	Highlights of activities included in Monthly Report	1) Section head Bureau director 2) Editor/ admin. asst. Section heads Bureau director	Expose research activities to Department	The Department and all regional offices are informed of practical advances in technology and research activities
Statistical Activities: 1) Consultation 2) Training	1) Status log of consultation projects Statistical consultation request forms 2) Course evaluations	1&2) Client's needs are satisfied 2) Training is appropriate for particular program area	1) Personnel/ resource justification 2) To evaluate quality of training provided	1) Log is updated when new projects/ activities are added 2) After completion of each training program	1&2) Assoc. stat. Section head Bureau director Division director 2) Training bureau and client	1) Keep track of progress and workload 2) Feedback on training quality/client satisfaction	Statistical consultant is available to entire Department Statistical training is tailored to needs of particular program area
Coordination of Department-Sponsored Research (including in-house & contract research, NCHRP, TCRP, pooled funds, etc.)	1) Semi-Annual Progress and Expenditure Report 2) Research solicitations 3) Briefing books for research advisory panels 4) "Regular" communication with contract management project manager, NCHRP, TCRP, FHWA . . .	1) Under budget and on time 2&3) To formulate research program that addresses Department's needs and meets needs of client; Research is addressed through the appropriate program (in-house, contract, or Federal, etc.) 4) Documents are prepared on time, deadlines are met, effective liaison is maintained	1) Status of research program and projects reported 2) Circulated to stakeholders also published in newsletter 3) Status of research program reported; Recommendation and background information on each research suggestion is provided 4) Varies from formal memos to phone call and informal meetings	1) Every 6 months 2) Annual 3) Annual 4) As needed, varies from weekly to annual	1) FHWA 2) All stakeholders, including all department programs, academia, and contractors 3) Bureau director Division director Members of appropriate research panel that the briefing book is prepared for 4) Important activities are included in Monthly Report	1) Communi- cate progress, keep projects on schedule and under budget 2) Client needs are addressed 3) Keep track of each research solicitation received and its disposition 4) Each program and funding source is effectively utilized, research projects are initiated in a timely manner, and progress of each project is monitored	This function provides focus to research activities of Department It provides a process by which Department's strategic research needs are effectively defined, prioritized, and implemented

TRANSPORTATION RESEARCH & DEVELOPMENT BUREAU—TECHNOLOGY TRANSFER (New York)

Function	Performance Measure	Goal	Used: How	Used: When	Used: By Whom	Used: For What	Benefits/ Comments
Support Functions (incl. computer/storeroom/vehicles/ electronic lab)	All support activities progressing smoothly (no complaints)	These functions support other functions to ensure that the program is running smoothly Future needs are determined and planned Inventories are maintained Necessary paperwork is prepared in a timely manner Vehicles and equipment are safe and in working order	To keep track of activities	Noteworthy activities are included in Monthly Report	Section head Bureau director (if noteworthy)	Monitor work done in these areas	These areas virtually run themselves; section head only gets involved if there is a problem

UTAH DOT

PROJECT DEVELOPMENT GROUP
PERFORMANCE MEASURES

Vision: *We are one team pulling together to support the Regions & Divisions in developing quality transportation projects.*

Mission: *We support the development of quality projects which are environmentally sensitive, biddable, and constructable, providing superior performance to assure customer satisfaction.*

Key Business Area/Function	Customer Expectations	Desired Outcome	Input Measure	Output Measure	Efficiency Measure	Outcome Measure
Division/Section: Research Division **Mission:** *Our mission is to provide useful and timely information about new technologies, products, or procedures, to improve the operations, safety, efficiency, and cost effectiveness of our transportation systems.*						
Identify & prioritize research topics	The most pressing problem facing the Department will be funded.	- A work program aimed at the Strategic Goals of UDOT - Satisfied stakeholders	- Person hours expended on UTRAC process - Cost of process	- Number of regions/divisions present - Total number of ideas from workshop - Number of topics funded	- % of budget expended on the UTRAC process - % divisions, regions, & other organizations at workshop - % of problems funded from list	- Through region & division visits, identify if the most pressing issues are receiving prioritization, and getting funded as studies through the process. - Conduct surveys

Key Business Area/Function	Customer Expectations	Desired Outcome	Input Measure	Output Measure	Efficiency Measure	Outcome Measure
Develop projects & work plans	The plan developed has a good chance to find a solution to the problem and keeps the stakeholders involved throughout.	A complete & well-defined work plan for each project that has been reviewed & approved by the problem champions & TAC	- Number of studies programmed	- Number of work plans completed - All tasks included in the work plan	- % of work plans completed - % of needed tasks included in the plan	Interview TACs verbally as well as through surveys to evaluate the quality & completeness of work plans.
Conduct and/or manage projects	Study objectives achieved. The recommendations & findings coincide with the problem statement & plan.	Contribute to knowledge base. Recommend improved specifications, policies, procedures, etc. Deliver design tools, expert systems, lab tests, software packages, etc.	- Cost to manage the study - Total cost to complete study - Months programmed to complete each study	- Number of studies completed - Months needed to complete each study	- % studies completed vs. planned - % of funding to manage each study - % of funding required to complete each study - % of TAC meetings held & members there - % of final reports - % of total research budget expended	Conduct exit surveys through TACs to evaluate the quality & completeness of study deliverables. Region/division visits will be used to evaluate the value of completed studies. Gather information on benefits achieved.
Implement findings from research studies conducted by UDOT & other agencies	Timely delivery of quality research products in a useable form.	Research findings are put into practice, and function as required.	- Implementation expenditures - Number of initiatives planned	- Number of initiatives implemented - Estimated benefits in the form of cost savings, enhanced movement of people or goods, or improved safety.	- % of implementaton plans prepared - % of planned initiatives completed - % of funds expended on implementation - Benefit/cost estimate for the program	Track number & worth of study deliverables implemented into UDOT operations through surveys & visits.
Information exchange, technology transfer & peer exchange	Timely & appropriate products representing state-of-the-art information.	Literature & report distribution, workshops, training, presentations, InfoX sessions, TransX sessions & newsletters	- Number of items requested or identified as action items - Cost of info exchange - Number of complaints received	- Number of action items completed - Number of complaints resolved	- % of newsletters published vs. planned - % of action items completed - % of complaints resolved - % of infoX & TransX sessions held	Estimate value to UDOT personnel & operations of information exchange activities through questionnaires and visits.
New products evaluation	Fair & timely review of all submitted products by responsible functional groups & the New Products Evaluation Panel (NPEP).	Current Product Acceptability Listing (PAL), technical evaluation of the products, action plan for the use of each product.	- Number of products submitted for review - Cost for new product reviews - Number of complaints	- Number of new products reviewed - Number of complaints resolved	- % of submitted products reviewed - % new product program of total budget - % of complaints resolved	Calculate value of new products implemented into practice. Survey stakeholders related to the new products review process.
Experimental features	Timely & accurate evaluation with appropriate recommendations based on sound data.	A technical & economic evaluation resulting in a recommended action plan for use of the feature.	- Number of planned experimental features - Cost of establishing each feature	- Number of features evaluated - Benefits of the program	- % of planned feature completed - Benefit/cost of experimental features program	Solicit from appropriate region personnel estimates on value & quality of experimental features.

Key Business Area/Function	Customer Expectations	Desired Outcome	Input Measure	Output Measure	Efficiency Measure	Outcome Measure
Develop funding sources to maximize project & program value	Projects & programs funded at an adequate level. New money will be aimed at UDOT's strategic goals.	Enhanced existing budget to solve as many of UDOT's goals as possible.	- Cost of submitting external proposals	- New money added to budget - Hours spent by staff members securing funds	- Benefit/cost of submitting proposals for new funding - % of total budget spent on external proposals	Estimate value of the studies funded by external sources. National partners and funding agencies will be used as a data source.

PERFORMANCE MEASURES FOR THE VIRGINIA TRANSPORTATION RESEARCH COUNCIL TECHNOLOGY STRATEGIC OUTCOME AREA

Customer Satisfaction: The extent to which VTRC's customers are satisfied by the choice of projects undertaken by the staff, timeliness and accuracy of our research, applicability of research to customer problems, responsiveness to requests for technical assistance, and quality of VTRC products.

Performance Indicators:

- Customer Satisfaction Surveys
- Customer Anecdotes
- No. of Requests for Library Services
- No. of Requests for VTRC Publications

Innovations Implemented: The extent to which VTRC's research program has led to beneficial changes in improved specifications, extended service life of facilities, operational efficiencies, cost savings, improved safety, etc. for VDOT and the traveling public.

Performance Indicators:

- No. of Specifications Revised
- No. of VTRC Recommendations Implemented
- No. of New Methodologies Implemented
- No. of Facilities with Extended Service Live
- No. of New Products Evaluated and Implemented
- No. of Legislative or Policy Changes
- Dollar Value of Costs Saved/Avoided

Advancing the State of the Art in Transportation: The extent to which the VTRC is influential in advancing the state of the art and practice at the state and national level.

Performance Indicators:

- No. of Referred Publications
- No. of Papers/Presentations at National Conferences

- No. of Committee Memberships/Chairmanships on Technical Committees
- No. of VTRC Reports Published
- No. of "Hits" on VTRC Web Site

Develop Tomorrow's Transportation Professionals: The extent to which the VTRC's Technology Transfer Program and Graduate Research Assistantship Program produces qualified transportation professionals to serve the needs of the Commonwealth.

Performance Indicators:

- No. of Graduating Transportation Students Supported by VTRC at UVa and VPI
- No. of Short Courses/Workshops Taught by VTRC staff
- No. of LTAP-Sponsored Courses
- No. of LTAP Workshop Participants

APPENDIX D

New York Benefit-Cost Procedure

FORM 1: BENEFIT CHECKLIST

Benefit Areas and Specific Benefits	Net Benefits
CAPITAL PROGRAM	
Labor	
Materials	
Equipment	
Service Life	
Contract Duration	
Other: _____	
Other: _____	
OPERATING PROGRAM	
Maintenance Frequency	
Maintenance Equipment	
Maintenance Labor	
Maintenance Materials	
Design Procedures	
Other: _____	
Other: _____	
SAFETY AND USERS	
Accident Frequency	
Accident Severity	
Other:	
Other:	
Travel-Time Savings	
Vehicle-Operating Costs	
Other: _____	
Other: _____	

NOTE: Additional benefits may exist and should be added if monetary amounts are significant.

APPENDIX B
GUIDELINES FOR CALCULATING BENEFITS OF RESEARCH

This appendix presents guidelines for analysis and calculation of benefits from completed and proposed research projects, including the life-cycle and present-worth methods used by the Transportation Research and Development (TR&D) Bureau. These procedures were designed to estimate monetary values in three areas.

1. Capital Program Benefits: primarily savings in the Department's out-of-house expenses (such as contract construction), discussed here in terms of 1) labor, 2) materials, 3) equipment, 4) longer service life, and 5) contract duration.

2. Operating Program Benefits: savings affecting the Department's in-house expenses, discussed here in terms of 1) maintenance, and 2) design procedures.

3. Safety and User Benefits: savings associated with improved highway safety (in terms of accident frequency and severity) and both reduced travel time and vehicle operating/maintenance costs for users.

This benefit-analysis procedure involves calculating net annual benefits for each project within these three benefit areas, and converting net annual totals to five-year life-cycle, present-worth costs. "Net benefits" are defined here as aggregate benefits, minus all resulting increased costs of implementation (such as labor, materials, or equipment). When calculating research performance measures, benefits may be totaled into one single figure. When calculating benefits for proposed projects, benefits should be reported for each of these three benefit areas on a project-by-project basis. Benefits of research may be calculated using the following four-step procedure.

Step 1: Identify Benefits

This involves matching a project's benefits into one or more of these three broad benefits areas, as well as against specific benefits within each area. This step is important to avoid overlooking anything that could significantly affect the results. The checklist in Form 1 is used, listing the specific benefits found in each area. (Additional information about these specific benefits is given later in these guidelines.) Benefits analyses are often specific to a particular project, and benefits may result other than those listed in Form 1. If such benefits are found, they should be included in the appropriate area and listed under "other." This step also includes identifying any increased costs associated with implementing research recommendations. Increased costs should be calculated annually by benefit area, and subtracted from the benefits before converting the area totals to present-worth values.

Step 2: Define Target Area

This involves identifying specific geographic areas of New York State that will be affected by a project's results. This is important, especially in terms of regional areas, because certain costs vary significantly according to geographic location. "Regions" are defined as follows for benefit calculations:

1. Upstate: all locations except Bronx, Kings, Nassau, New York, Queens, Richmond, and Suffolk Counties.
2. Metropolitan: the seven counties just listed in New York City and on Long Island.
3. Statewide, all counties in New York State.

Step 3: Calculate Benefits

In this step, benefits are calculated for the three benefit areas. Two sample problems are provided as examples at the end of these guidelines (pp. 42–44), with results summed as annual dollar savings. If annual benefits for a specific project may change from year to year they should be calculated annually for a five-year period, as in the case of such cumulative benefits as safety-related savings. Assumptions may be necessary during benefit calculation, typically when needed data either do not exist or are extremely difficult to obtain. Resulting assumptions require good engineering judgment, and should be clearly identified as assumptions to avoid confusion when calculations are reviewed.

Step 4: Calculate Present-Worth Life-Cycle Cost

This involves converting annual benefits to a present-worth value, and requires two assumptions:

1. Life Cycle: TR&D bases calculations on benefits resulting during only the first five years of implementation.
2. Interest Rate: a 4-percent rate is used for all benefit calculations.

There are two different methods for calculating present worth, depending on type of annual benefits. The method shown on Form 3 should be used if annual benefits are *uniform* during the five-year period. The method shown on Form 3 should be used when annual benefits are *non-uniform* throughout the five-year period. Sample calculations for both uniform and non-uniform present-worth calculations are provided at the end of these guidelines.

Capital Program Benefits

Labor

Labor benefits are applicable when a study's anticipated results can reduce labor cost of a capital project. These benefits should be determined based on current wage rates and fringe-benefit percentages. It is important to remember that labor rates in New York State can vary substantially by county—thus, one must define the area within the state to be affected by the project. The following assumptions are used for all labor-benefit calculations:

1. All labor benefits having statewide effects are calculated using statewide average labor rates.
2. Labor benefits that are regional or occur in New York City are calculated using representative labor rates for that area.

FORM 2: PRESENT-WORTH LIFE-CYCLE-COST WORKSHEET FOR UNIFORM ANNUAL BENEFITS

Benefit Area	Annual Benefit ($/year)	A/P Factor	Present Worth ($)
Capital Program	_____	× 4.4518 =	_____
Operational Program	_____	× 4.4518 =	_____
Salary and Users	_____	× 4.4518 =	_____

Data: I = interest rate = 4.0%
N = life-cycle = 5 years
A/P Factor = converts annual payments to present-worth value for a given life-cycle and interest rate

$$A/P\ (4.0\%, 5) = ([1 + I]^N) - 1/I \times (1 + I)^N$$
$$= ([1.04]^5) - 1/0.04 \times (1.04)^5$$
$$= 4.4518$$

NOTE: This form should be used only when annual benefits within each area do not change from year to year.

FORM 3: PRESENT-WORTH LIFE-CYCLE-COST WORKSHEET FOR NON-UNIFORM ANNUAL BENEFITS

Benefit Area and Year	Annual Benefit ($/year)	F/P	Present Worth ($)	Group Totals
CAPITAL				
1	_____	× 0.962 = $	_____	
2	_____	× 0.925 = $	_____	
3	_____	× 0.889 = $	_____	
4	_____	× 0.855 = $	_____	
5	_____	× 0.822 = $	_____	$_____
OPERATING				
1	_____	× 0.962 = $	_____	
2	_____	× 0.925 = $	_____	
3	_____	× 0.889 = $	_____	
4	_____	× 0.855 = $	_____	
5	_____	× 0.822 = $	_____	$_____
SAFETY				
1	_____	× 0.962 = $	_____	
2	_____	× 0.925 = $	_____	
3	_____	× 0.889 = $	_____	
4	_____	× 0.855 = $	_____	
5	_____	× 0.822 = $	_____	$_____
USERS				
1	_____	× 0.962 = $	_____	
2	_____	× 0.925 = $	_____	
3	_____	× 0.889 = $	_____	
4	_____	× 0.855 = $	_____	
5	_____	× 0.822 = $	_____	$_____

Data: I = interest rate = 4.0%
N = life-cycle = 5 years
Factor = converts benefits to present-worth value for a given life-cycle and interest rate

$$Factor\ (4.0\%, N = 1 - 5) = 1/(1 + I)^N$$

Data Sources

Statewide Averages:
1. Main Office: Design Quality Assurance Bureau or Contract Management Bureau
2. Prevailing Wage Unit (718) 797-7731

Regional and NYC Rates:
1. Study proposals for regional projects from the Design Quality Assurance Bureau
2. New York City Controller's Unit: (212) 566-2170
3. Upstate rates: NYS Department of Labor: (518) 457-5589

Materials

Capital program benefits can result from reduction in materials quantities and/or costs, and should be based on current market values. Note that many sources of information concerning current materials costs may also provide information on associated labor costs—for example, the Weighted Average Bid Prices published semiannually by NYSDOT, providing information on materials quantities and costs at numerous locations statewide. Not only are materials covered, but also labor and equipment. When bid prices must be converted to pure materials costs, assume that materials are one-third of total cost.

Data Sources

1. BAMS (Bid Analysis and Management System): this is a NYSDOT computer system containing the most up-to-date records for all bid prices statewide. It is designed to call up any bid price or quantity on any contract dating back to 1984.
2. Weighted Average Bid Prices: Office of Engineering, NYSDOT, available in the TR&D Library.
3. Design Quality Assurance Bureau and Management Bureau.

Equipment

Benefits can be achieved through reducing or improving needed equipment, with calculations based on current operating and equipment costs.

Data Sources (publications available from the Construction Division)

1. Costs Reference Guide for Construction Equipment: Dataquest, Inc., San Jose, Calif.
2. Rental Rate Blue Books: Equipment Guidebook Co., Palo Alto, Calif.
3. Means Heavy Construction Cost Data: R.S. Means Co., Kingston, Mass.

Longer Service Life

Longer lives for capital-improvement projects are beneficial to the Department by reducing the need for reconstruction, redesign, and related operations. (No data sources are suggested, because service life is highly specific to a given project, material, or process, but references previously listed here may be helpful.)

Contract Duration

Reduction of contract duration and contract-time overruns provide significant monetary benefits for the Department's capital program, especially when lane-closure, lane-rental, and contract-completion-incentive clauses are included in a contract. (When contract-duration savings are a potential benefit for a project, the Construction Division should be consulted; other data sources listed here may also be pertinent in calculating this benefit.)

Operating Program Benefits

Maintenance

This is a major part of the Department's operating program, and may be identified as having four separate areas for purposes of identifying benefits: 1) frequency, 2) equipment, 3) labor, and 4) materials. Monetary benefits to the operating program may result from savings in any or all of these four separate areas. Less frequent maintenance provides savings to the operating program by allowing workers to conduct more maintenance activities in a given year. Reductions in maintenance equipment, labor, and materials needed to complete a given task also represent monetary benefits for the program. Increases in any of these items must be considered when calculating net benefits. Benefits in maintenance practices should be calculated using maintenance-cost figures provided by the Transportation Maintenance Division. Because most maintenance activities are conducted by Department forces, it may be inappropriate to use contract bid prices for benefit calculations, except where the Department has contracted for maintenance work.

Data Sources

1. For Department-conducted maintenance: Transportation Maintenance Division (main office and regions)
2. For contract maintenance: Weighted Average Bid Prices, Office of Engineering
3. BAMS (Bid Analysis and Management System)
4. Bridge Inventory System
5. Means Heavy Construction Cost Data, R.S. Means Co., Kingston, Mass.
6. Rental Rate Blue Books: Equipment Guidebook Co., Palo Alto, Calif.
7. Equipment Management Division

Design Procedures

Savings to the operating program can be calculated in terms of cost-effectiveness of the Department's design procedures, including anything from improved design methodologies to improved computer-aided-design packages.

Data Sources (NYSDOT Main Office)

1. Structures Design and Construction Division
2. Design Division (Main Office)
3. Engineering Automation Resource Support Group (Main Office)

Safety Benefits

Accident Frequency

Reduction in accident frequency results in significant savings for the Department. Among the benefits are less maintenance of damaged roadway facilities, fewer injuries and fatalities, and improved road operating characteristics. Although many sources exist for accident data and statistics, information from the New York State Department of Motor Vehicles (NYSDMV) and NYSDOT's Transportation Planning, Highway Safety, and Traffic Engineering Division (TPHSTE) should be used in calculating benefits accruing from reduced accident frequency. (Other sources can be helpful in some instances, but may not reflect typical New York State conditions.)

The TPHSTE Division continuously develops figures for accident-reduction percentages based on road improvements. They also have worksheets specifically designed for safety-related benefit calculations that provide an organized acceptable means for calculating these benefits.

Data Sources

1. Annual Accident Summaries, NYSDMV.
2. Long, G.H., and Watson, J.E. Highway Safety Improvement Program: Procedures and Techniques, NYSDOT Traffic Engineering and Safety Division, November 1989.
3. Accident Facts (published annually), National Safety Council

Accident Severity

Reduction of accident severity is another safety benefit for the Department, particularly in cases of liability against the state. By reducing severity, risk of personal injury is diminished as well as likelihood of legal action. Numerous sources exist for accident information, but statistics developed by NYSDMV and NYSDOT's TPHSTE Division are the most useful.

Data Sources

1. Annual Accident Summaries, NYSDMV.
2. Long, G.H., and Watson, J.E. Highway Safety Improvement Program: Procedures and Techniques, NYSDOT Traffic Engineering and Safety Division, November 1989.
3. Accident Facts (published annually), National Safety Council

User Benefits

Travel Time

These are benefits to system users, rather than monetary benefits to the Department. An interesting characteristic of user benefits is their magnitude compared to other benefits. Because user benefits affect the entire user public, they generally result in very high dollar values. Benefits affecting travel time and travel distance are very similar and often identical. Both depend on monetary values placed on the user's time, but this by itself is controversial because users generally value their time very specifically. Many past studies to develop personal dollar values have themselves differed substantially. Travel-time savings affect three major groups: commercial-vehicle operators, business travelers, and recreational travelers. Particular assumptions are specific to each type of travel-time benefit:

1. Commercial-Vehicle Operators: they often assume that any time-saving can be translated into additional output by the crew and perhaps the vehicle. These should be evaluated with caution, because it is often difficult (if not impossible) to reassign vehicles and crews to other operations.

2. Business-Travel Time: assumptions for business-travel time are similar to those for commercial operations, with these additional complications: 1) considerable business travel occurs during the road user's personal time and savings thus are not necessarily translated into extra output or work, and 2) people often work while traveling, so that their travel time is not always truly "lost" time.

3. Recreational-Travel Time: these values are expressed as a proportion of average-hourly wage rate on the assumption that willingness to pay for time savings is related to income measured by wage rate. The following table, prepared by the Institute of Transportation Engineers, lists recommended values for some in the United States, and should be used for all travel-time calculations:

Data Sources

1. Transportation Planning Handbook, Institute of Transportation Engineers, 1992.
2. A Manual of User Benefit Analysis of Highway and Transit Improvements, AASHTO, 1977.

Person-Time (Person-Hours)	In-Vehicle Time ($/Person-Hour)	Waiting/Walking Time* ($/Person-Hour)
Low Time Savings (0–5 min)		
Average Trips	0.21	5.85–7.80
Work Trips	0.48	5.85–7.80
Medium Time Savings (5–15 min)		
Average Trips	1.80	5.85–7.80
Work Trips	2.40	5.85–7.80
High Time Savings (over 15 min)		
Average Trips	3.90	–
Work Trips	3.90	–
Commercial Vehicles		
		Overall
Single Unit Truck (SUT)		7.00
3–S2 Design Vehicle		8.00

Depends on out-of-vehicle comfort and safety

Vehicle-Operating/Maintenance Cost

Improvement of road conditions or road-design policies may affect vehicle maintenance and operating costs. Again, these savings might seem fairly trivial, but can be substantial when the entire user public is concerned. Savings in vehicle operating/maintenance costs producing monetary benefits for the user generally include 1) fuel consumption, 2) lubricants, 3) vehicle maintenance (labor, parts), 4) capital consumption (depreciation), 5) interest on capital employed, 6) wages, and 7) overhead.

Data Sources

1. Transportation Planning Handbook, Institute of Transportation Engineers, 1992.
2. A Manual on User Benefit Analysis of Highway and Bus Transit Improvements, AASHTO, 1977.
3. Zaniewski, J.P., Butler, B.C., Cunningham, G., Elkins, G.E., and Paggi, M.S. Vehicle Operating Cost, Fuel Consumption, and Pavement Type and Condition Factors. Report FHWA/PL-82/001, Texas Research and Development Foundation (Austin), March 1982.
4. The Highway Design and Maintenance Standards Model. Baltimore: Published for the World Bank by the Johns Hopkins University Press, 1987.

SAMPLE PROBLEM 2

Step 1

1. Title: Corrugated-Metal Box-Culverts (CMBC)
2. Description: Calculate potential benefits from using CMBCs (at least 22-ft width) instead of current precast-concrete box-culverts. Assume Department will continue using box-culverts at a rate equal to current use. This benefit analysis can be classified as "capital" based on Form 1.

Step 2 Statewide target area

Step 3

1. Assumptions:

1/3 of bid price represents materials
For CMBC, no costs are available for labor and equipment; although probably less than for precast concrete, they are assumed equal
4% interest rate
5-year life cycle
Design lives are equal

2. From BAMS and Weighted Average Bid Price book:

Two box-culvert items used in previous years (representative of previous annual use)

Item 1 (22' × 8')	Item 2 (22' × 8')
Avg bid price: $1600/lf	Avg bid price: $1719/lf
Amount placed: 57 lf	Amount placed: 40 lf

3. Calculate weighted avg bid price

[(57 lf) ($1600/lf) + (40 lf) ($1719/lf)]/97/lf = $1650/lf

4. Corrected weighted bid price for labor

= ($1650/lf) (1/3) = $550/lf for precast concrete

5. Department's annual materials cost

= ($550/lf) (97 lf) = $53,350/yr

6. From Corrugated Metal Box Culvert Mfg. Co.:

Approx materials cost = $330/lf

7. Potential Department cost using CMBC

= ($330/lf) (97 lf) = $32,010/yr

8. Annual savings: precast cost – CMBC cost
= $53,350 – $32,010 = $21,340/yr

Step 4 Using Form 2, because annual benefits are the same:

RESULTING BENEFITS = $97,730

APPENDIX E

Utah Customer Feedback Report

RESEARCH & DEVELOPMENT PROJECT
CUSTOMER FEEDBACK

Projects completed by the Research Division during FY 1996 and 1997 were given a letter grade by champions of the studies. The personnel grading the projects were asked to use the following criteria:

Grade	Description
A	Major impact—Revised operations
B	Significant impact—Improved operations
C	Contributed to state-of-the-art
D	Unclear or contradicting findings—More study needed
E	Major tasks not completed—Objectives not met

During this two-year period 39 projects were programmed. Four of the projects were carried over into the next fiscal year or discontinued, and were not graded. Following are the results of the process.

Grade	Number
A	14
B	15
C	6
D	-0-
E	-0-
discontinued	4

A grade point average for the program is estimated at 3.2 using this method.

This evaluation has proven to be useful to the program managers. An indication of the value of projects by type can be obtained. For example, during this time period projects related to materials studies received somewhat lower ratings than other types of studies. The performance of the study investigators is also easily evaluated. Changes in the program can be implemented, such as how the projects are awarded, funding by project type, training requirements for personnel, etc.

APPENDIX F

Utah Benefit-Cost Report

Benefits of Research & Development Projects

FY 1995, 96, and 97
May 2000

Project	Cost-$k	Benefit-$k	Benefit/Cost
Bridge Reinforcement with Composites	$160	$900	5.6
Composite Retrofit-Pushover Testing	$199	$4,000	20.1
Big Game Crossing	$49	$120	2.4
Teleconferencing at UDOT	$31	$200	6.5
Arcview Applications	$37	tbd	—
Management of Maintenance Features	$69	$500	7.2
Yellow–Green Safety Clothing	$8	tbd	—
Salt Detention Basins	$34	$4,500	132.4
Optimization of Traffic Signals	$19	$75	3.9
Open Graded Surface Course	$33	$120	3.6
Wetland Mitigation Banking	$164	$800	4.9
Drilled Shaft Resistance	$31	$600	19.3
Liquifaction Screening	$47	$900	19.1
CalTrans Seismic Review	$12	$750	62.5
Customer Satisfaction	$113	tbd	—
PM-10 Air Quality Study	$44	tbd	—
Solvent Based Paint Replacement	$43	tbd	—
Epoxy Bridge Deck Overlay	$20	$320	16.0
Structure I-Beam Access Deterrent	$25	$110	4.4
Pavement Marking Test Section	$60	$250	4.2
Delineator Posts Buttons vs. Tape	$15	$40	2.7
Concrete Patching Evaluation	$75	$900	12.0
Totals	$1,288	$15,085	11.7

tbd = to be determined.

APPENDIX G

Example Performance Measure Items from RTCC

TABLE A-1
ILLUSTRATIVE SEMI-HYPOTHETICAL EXAMPLES OF CHARACTERIZATION OF R&T FOR PERFORMANCE MEASUREMENT PURPOSES

Category of Research	Project	Direct Customer	Primary Stakeholders	Goal or Explicit Output	Intermediate Outcome	Impact on Transportation System Users
Operational	Advanced ATC Concepts (FAA)	Internal-Opn'l Offices	Airlines, airports, travelers	Improved opn'l concepts and tech.	More effective agency investment and performance	Improved capacity and reduced congestion
Regulatory	Fuel cell propulsion (USCG)	Internal-Opn'l Offices	Mariners	Improved technology	Better, more efficient mission performance	Improved marine safety; reduced govt. cost
	Tests & analyses for crashworthiness standards (NHTSA)	Internal-Regulatory Office	Automobile industry, motoring public	Guidance for regulation actions	More effective and efficient regulations and standards	Improved highway safety
	Truck dynamics studies (FHWA/OMC)	Internal-Regulatory Office	Trucking industry, motoring public	Understanding of truck dynamics	More effective and efficient regulations and standards	Improved highway safety
Policy	Cost-benefit of OPA Regulations (USCG)	OMB, Internal-Regulatory Office	Oil industry, shippers, public	Improved practices and equipment	More effective and efficient regulations and standards	Less environmental damage from oil spills
	HS Rail commercial feasibility (FRA)	FRA/DOT policy; Congress (ISTEA)	States, traveling public, equipment suppliers	Economic assessment	R&D and investment decisions	More effective investment of public funds
National Needs	Trans. impacts of telecommuting (OST)	Congress (mandate)	Workers, firms, telecomm. industry	Impact assessment	Federal policy decisions	Improved transportation planning
	ITS architecture and standards (FHWA)	State & local transp. authorities	Suppliers, hwy. users, Congress	Guidance for investments	More cost-effective ITS deployment	Improved mobility and system capacity and safety
	Maglev technology (FRA)	Congress (ISTEA)	Suppliers, states	Evaluation of tech. feasibility and cost	Federal, state, and private investments	Beneficial transportation innovation
	Alt. fuel buses (FTA)	Transit system operators	Suppliers, EPA	Evaluation of tech. feasibility and cost	Federal, state, and private investments	More effective investment of public funds

APPENDIX H

Applications of Performance Measurements to Various Phases of Research from Canadian Synthesis of Practice

APPLICABILITY OF EVALUATION METHODS TO TRANSPORTATION R&D			
Method	At the beginning—when deciding what R&D to conduct	During the R&D process	At the end—after the R&D has been completed
Expert opinion	A good supplement to user/client opinion (to provide them with information) and performance indicators (to value the indicators). Not particularly useful on its own.	Good way to monitor the technical progress of the R&D	Good way of obtaining high quality information regarding the quality and potential usefulness and impacts of the R&D, with relatively little effort.
User/client opinion	The best method for deciding what R&D to conduct, especially when supplemented by expert opinion or information from performance indicators.	Limited usefulness.	A fairly easy way of finding out if the R&D has been/is likely to be useful; but the results are more useful for program management purposes than reporting purposes.
Cost-benefit methods	Can be a useful way of prioritizing potential projects. Also, forces decision makers to think about the critical success factors. However, fairly time consuming.	Can be useful for monitoring if a cost/benefit analysis has been done at the beginning.	The best way of documenting the impacts of R&D whose benefits can be identified and quantified. Assumptions need to be clearly documented and, in general, the most challenging method to use correctly.
Case studies	Not useful	Not useful	The best way of documenting the impacts of R&D whose benefits can be identified but not quantified. Not particularly challenging.
Performance indicators	A good supplement to user/client opinion.	Can be useful in structuring the monitoring process.	Limited usefulness.

Transportation Association of Canada

THE TRANSPORTATION RESEARCH BOARD is a unit of the National Research Council, a private, nonprofit institution that provides independent advice on scientific and technical issues under a congressional charter. The Research Council is the principal operating arm of the National Academy of Sciences and the National Academy of Engineering.

The mission of the Transportation Research Board is to promote innovation and progress in transportation by stimulating and conducting research, facilitating the dissemination of information, and encouraging the implementation of research findings. The Board's varied activities annually draw on approximately 4,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation.

The National Academy of Sciences is a nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce Alberts is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encouraging education and research, and recognizes the superior achievements of engineers. Dr. William A.Wulf is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences, by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.